Course wrap-up

STAT 4710

December 5, 2023

Where we are

- ✓ **Unit 1:** Intro to modern data mining
- ✓ **Unit 2:** Tuning predictive models
- ✓ **Unit 3:** Regression-based methods
- ✓ **Unit 4:** Tree-based methods
- ✓ **Unit 5:** Deep learning

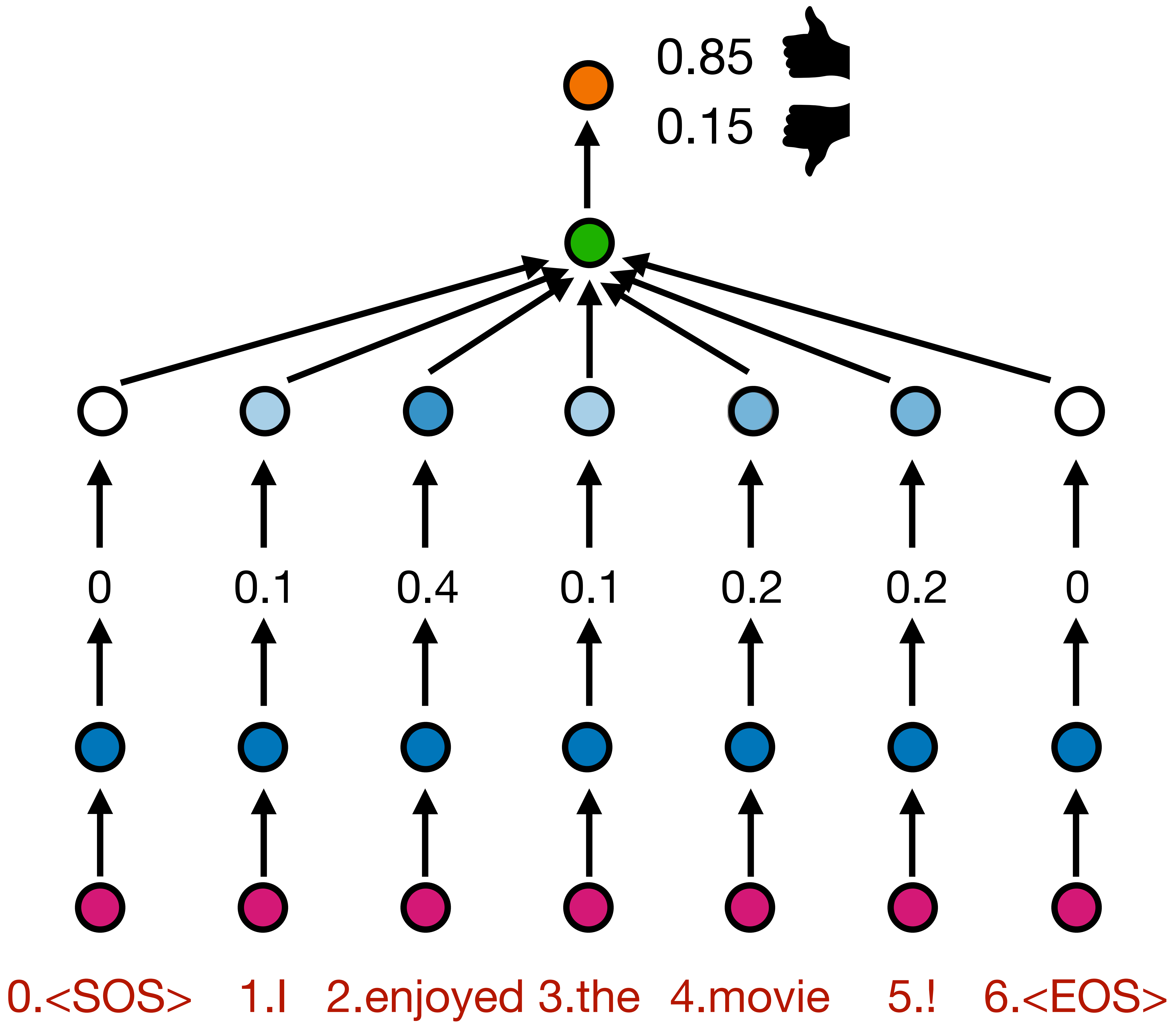
Where we are

- **Unit 1:** Intro to modern data mining
- **Unit 2:** Tuning predictive models
- **Unit 3:** Regression-based methods
- **Unit 4:** Tree-based methods
- **Unit 5:** Deep learning

Today's lecture:

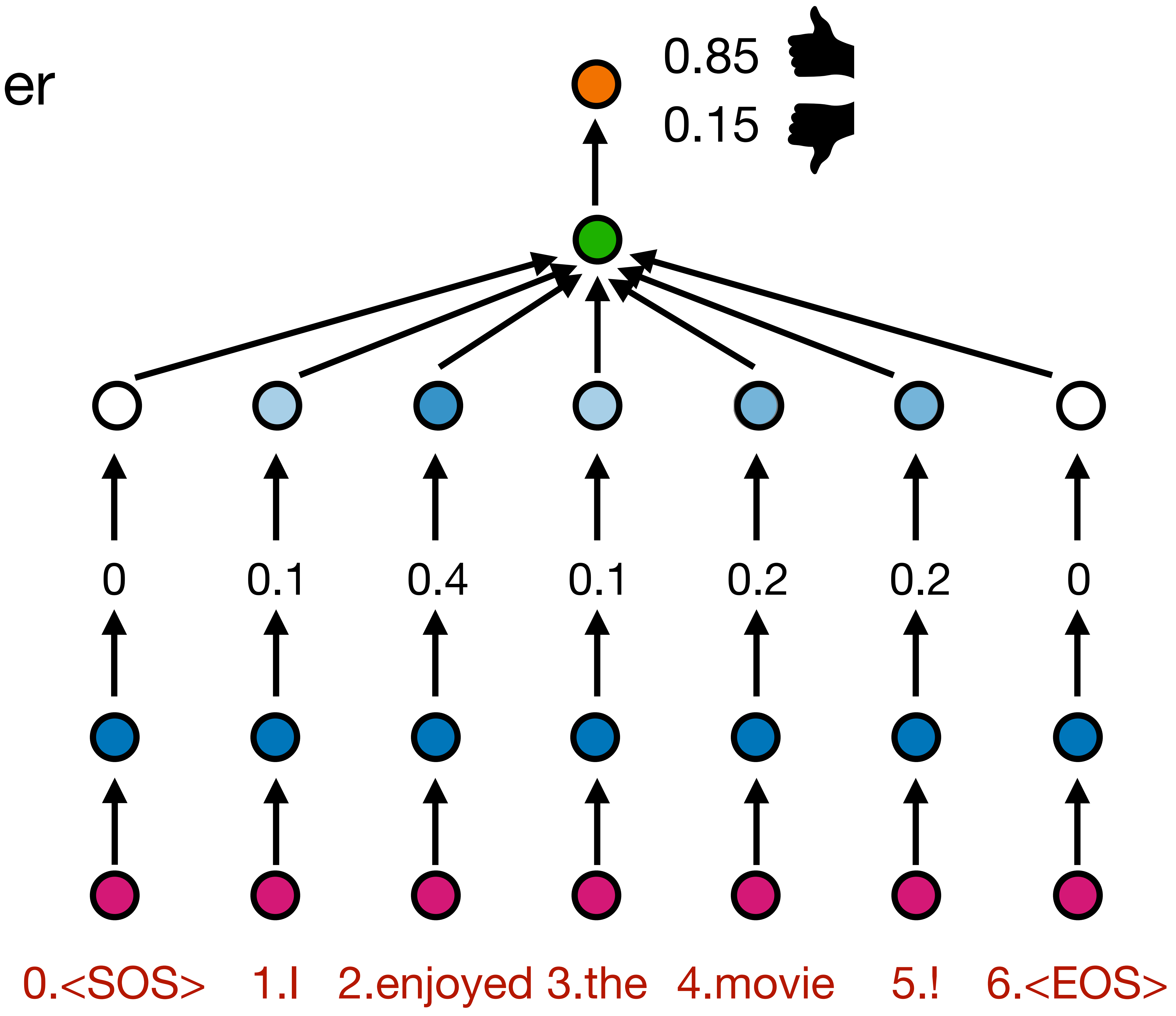
- Deep learning bonus material: Transformers beyond NLP.
- Looking back at STAT 4710
- Looking beyond STAT 4710

Transformers beyond NLP



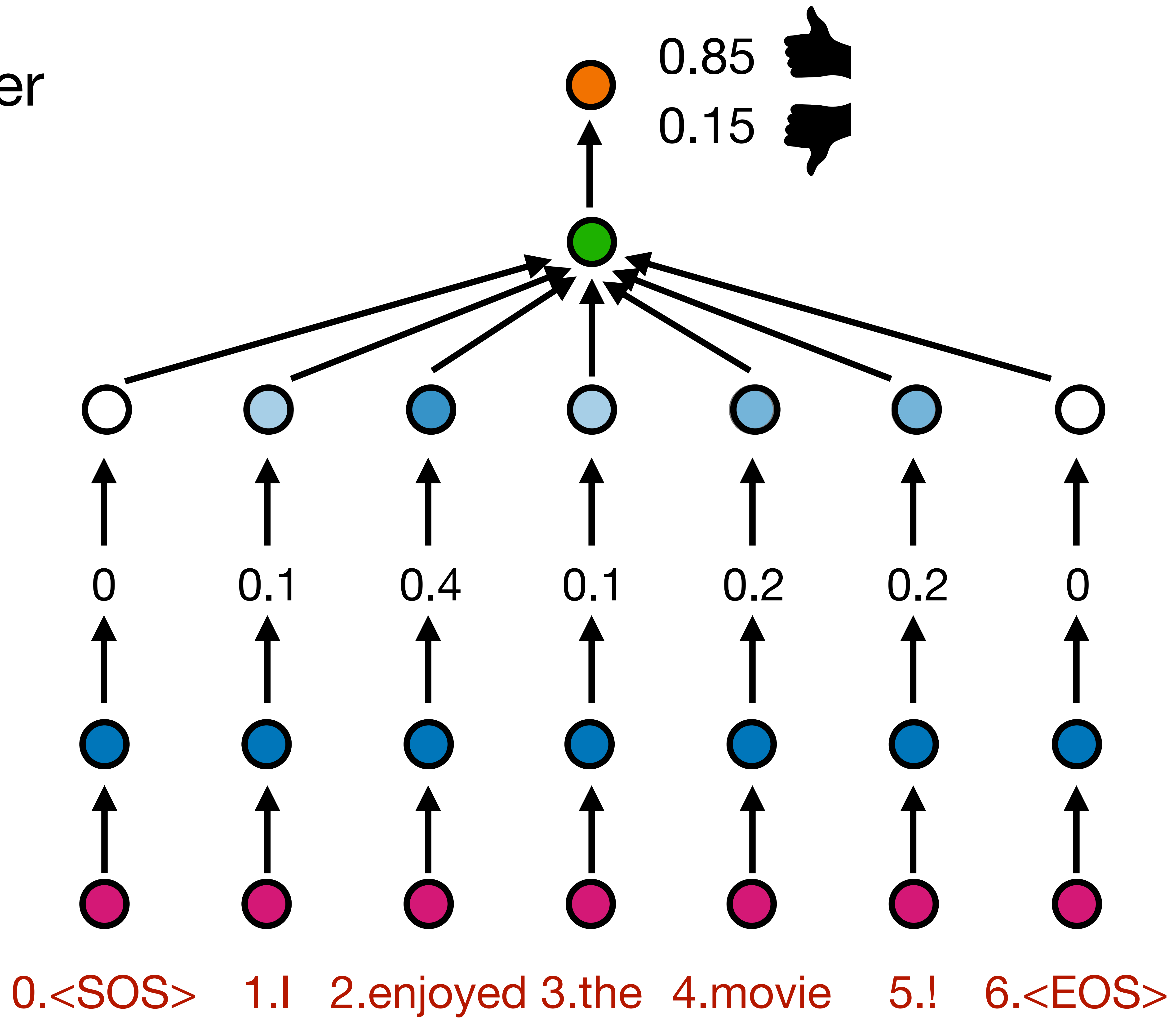
Transformers beyond NLP

- Transformers are being adapted to other domains, such as vision.



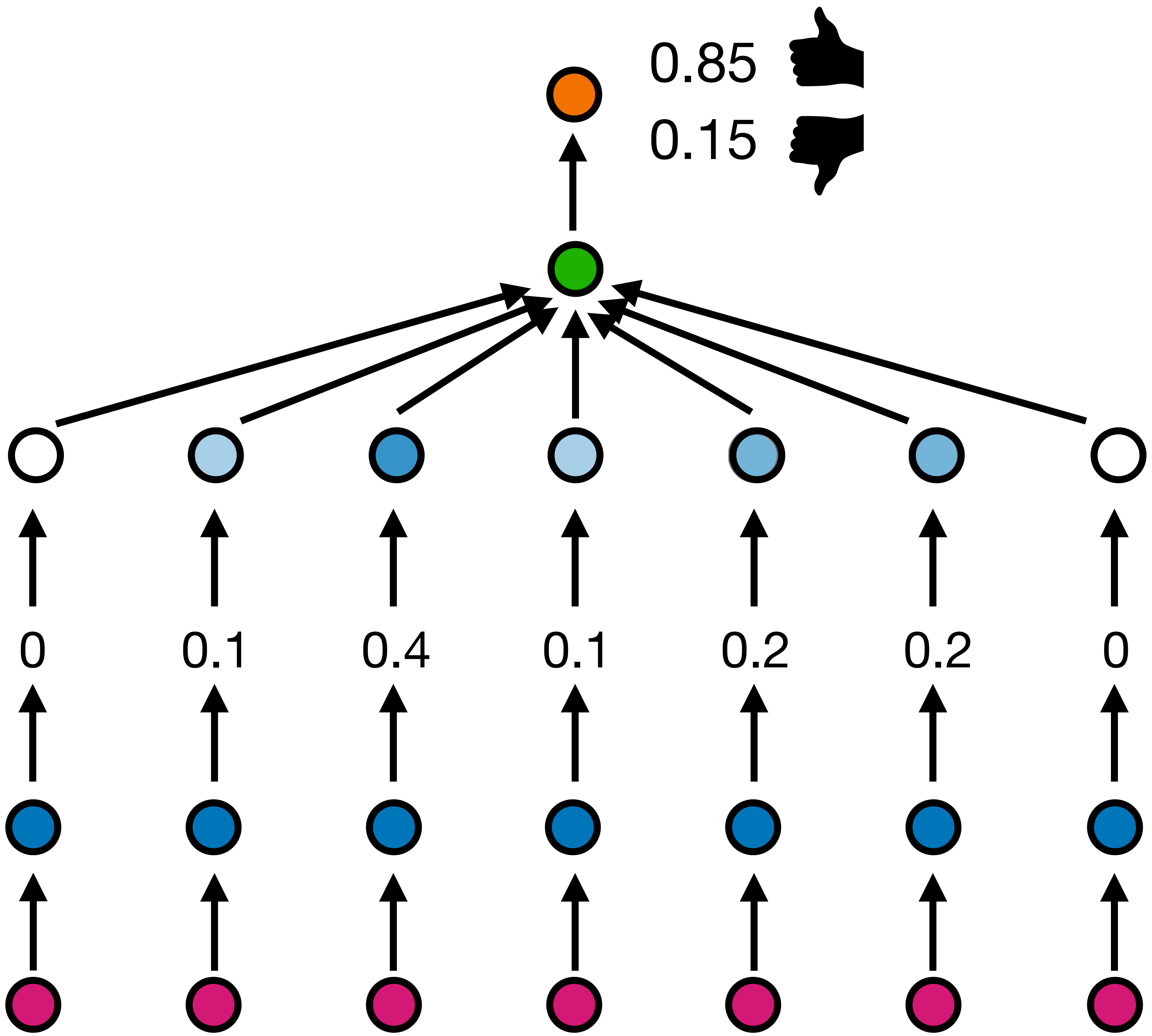
Transformers beyond NLP

- Transformers are being adapted to other domains, such as vision.
- Positionally-encoded image pieces are passed into an attention layer.



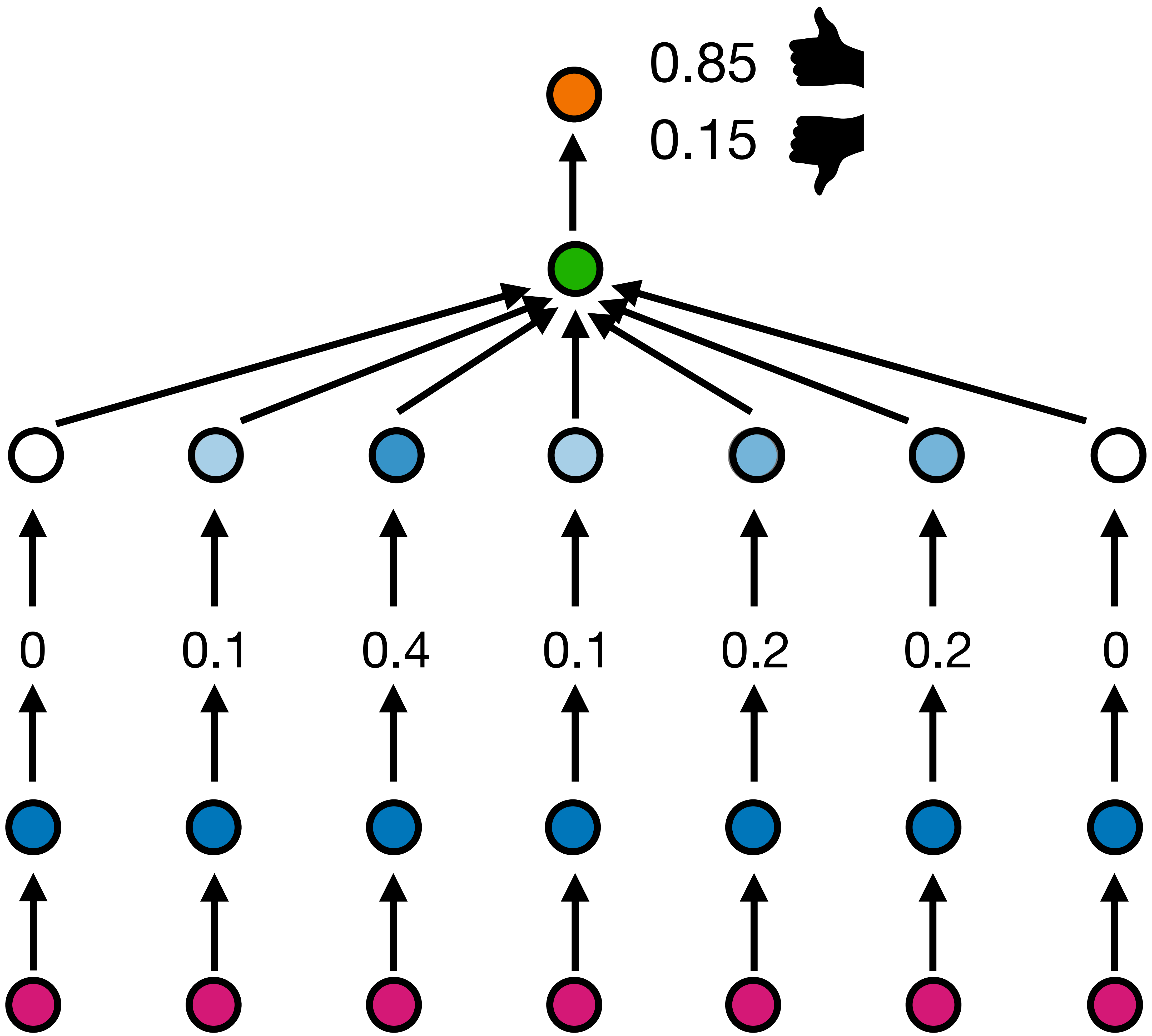
Transformers beyond NLP

- Transformers are being adapted to other domains, such as vision.
- Positionally-encoded image pieces are passed into an attention layer.



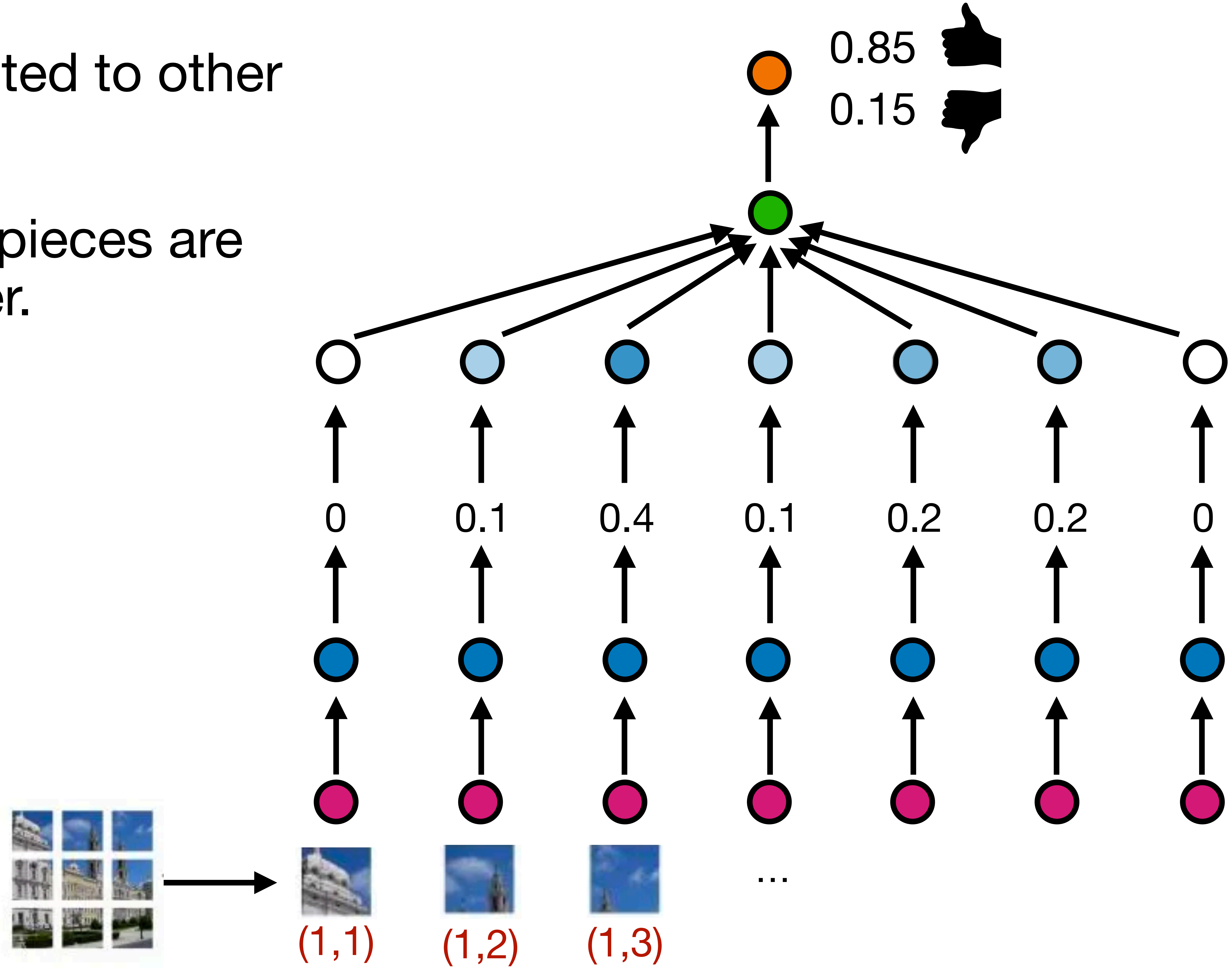
Transformers beyond NLP

- Transformers are being adapted to other domains, such as vision.
- Positionally-encoded image pieces are passed into an attention layer.



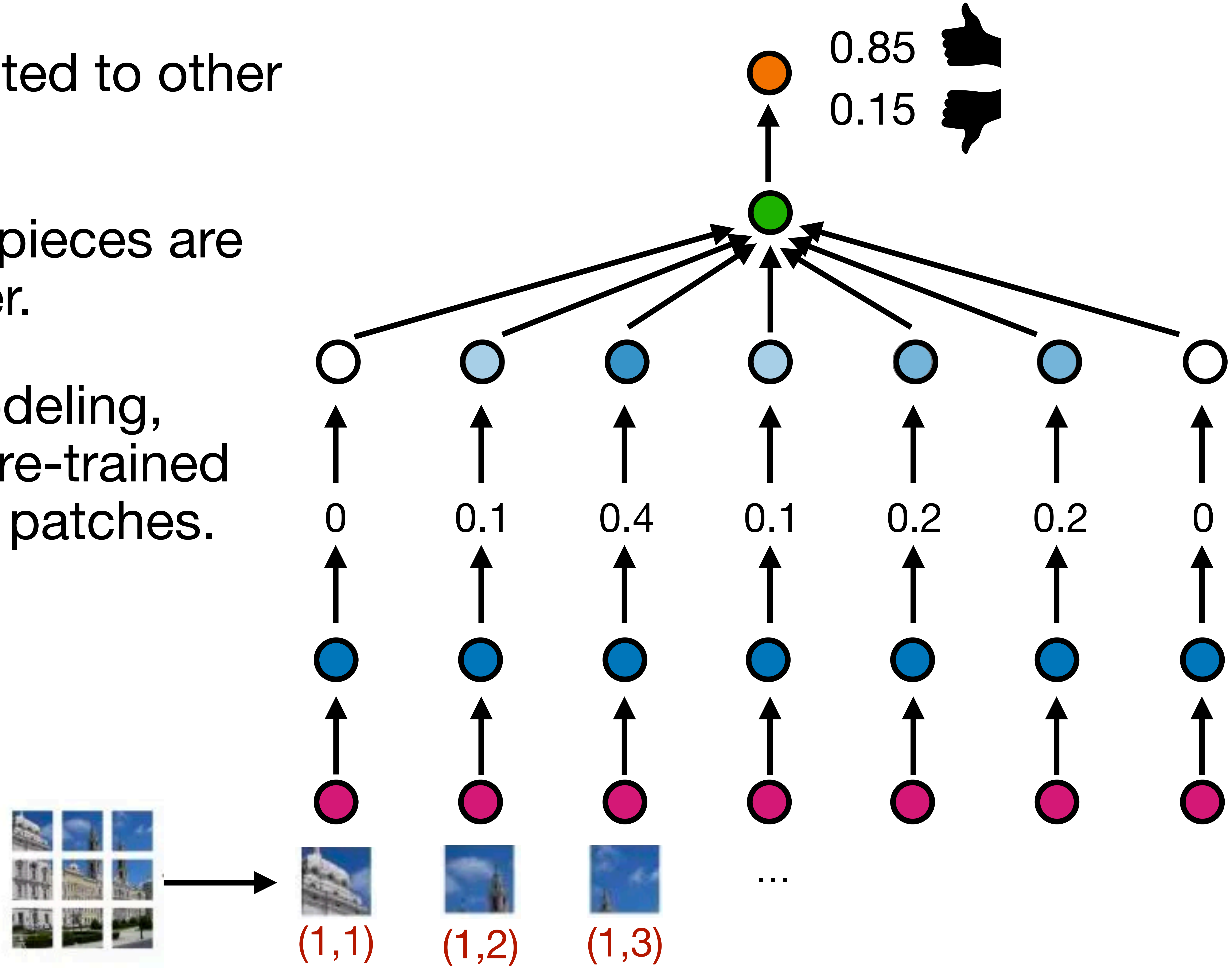
Transformers beyond NLP

- Transformers are being adapted to other domains, such as vision.
- Positionally-encoded image pieces are passed into an attention layer.



Transformers beyond NLP

- Transformers are being adapted to other domains, such as vision.
- Positionally-encoded image pieces are passed into an attention layer.
- Analogously to language-modeling, vision transformers can be pre-trained by predicting masked image patches.

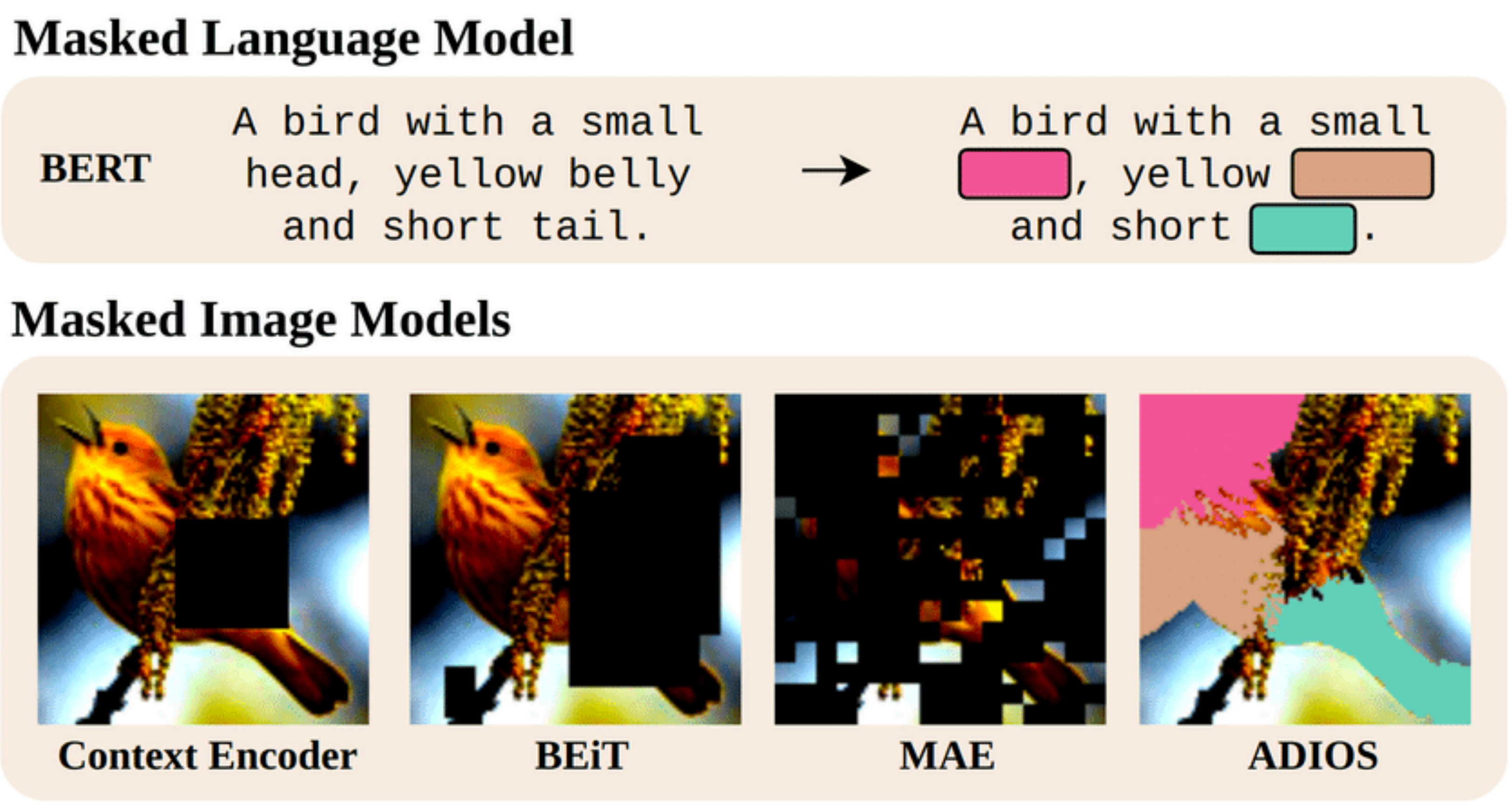


Transformers beyond NLP

- Transformers are being adapted to other domains, such as vision.
- Positionally-encoded image pieces are passed into an attention layer.
- Analogously to language-modeling, vision transformers can be pre-trained by predicting masked image patches.

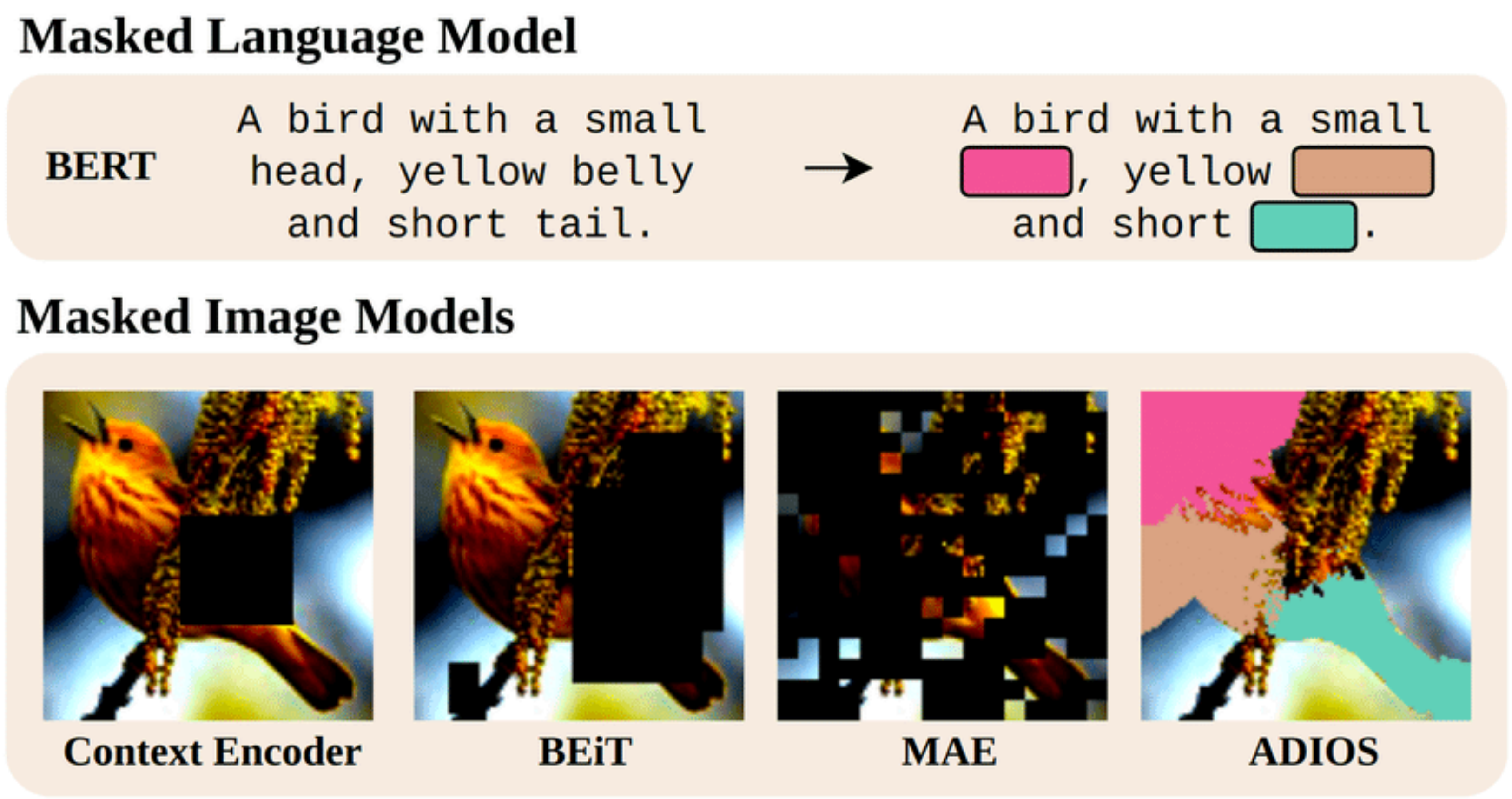
Transformers beyond NLP

- Transformers are being adapted to other domains, such as vision.
- Positionally-encoded image pieces are passed into an attention layer.
- Analogously to language-modeling, vision transformers can be pre-trained by predicting masked image patches.



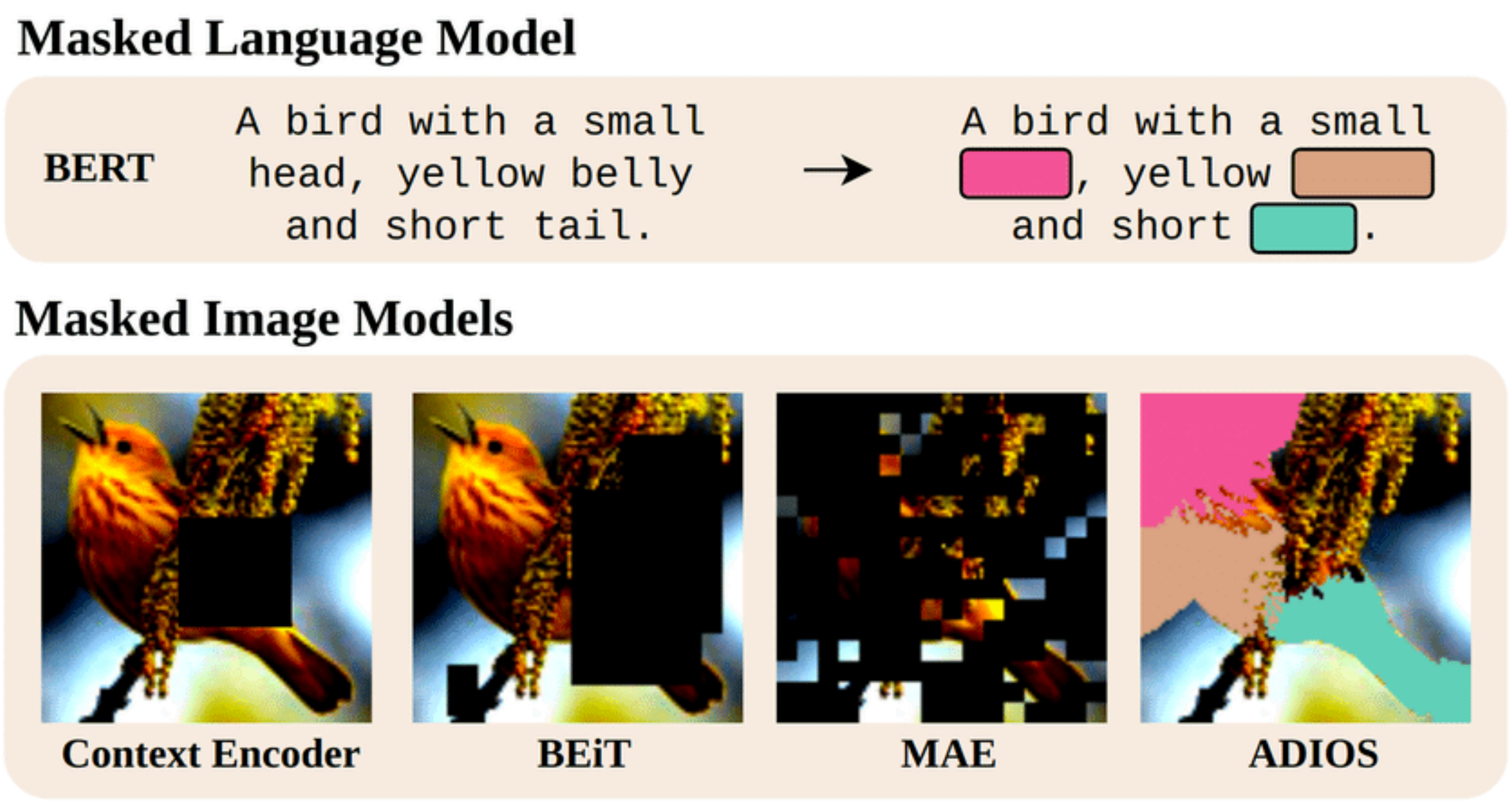
Transformers beyond NLP

- Transformers are being adapted to other domains, such as vision.
- Positionally-encoded image pieces are passed into an attention layer.
- Analogously to language-modeling, vision transformers can be pre-trained by predicting masked image patches.
- Multi-modal models are being trained, with text and images together.



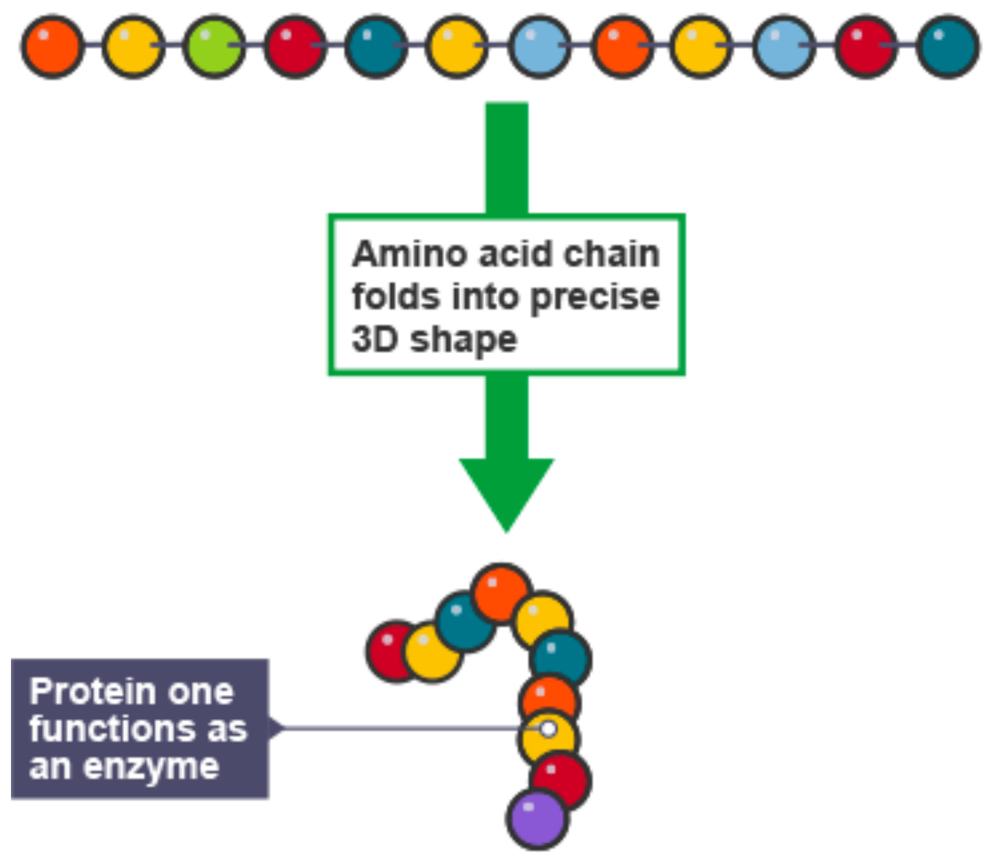
Transformers beyond NLP

- Transformers are being adapted to other domains, such as vision.
- Positionally-encoded image pieces are passed into an attention layer.
- Analogously to language-modeling, vision transformers can be pre-trained by predicting masked image patches.
- Multi-modal models are being trained, with text and images together.
- Transformers are also behind AlphaFold, breakthrough protein-folding model.



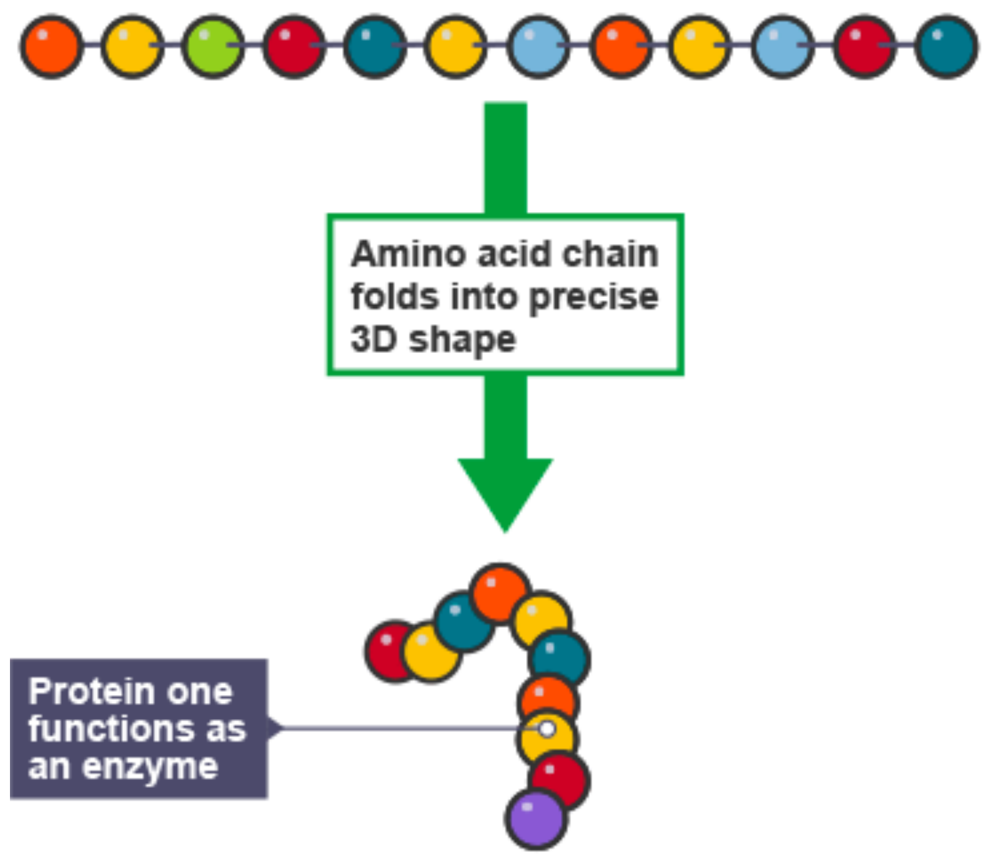
Transformers beyond NLP

- Transformers are being adapted to other domains, such as vision.
- Positionally-encoded image pieces are passed into an attention layer.
- Analogously to language-modeling, vision transformers can be pre-trained by predicting masked image patches.
- Multi-modal models are being trained, with text and images together.
- Transformers are also behind AlphaFold, breakthrough protein-folding model.



Transformers beyond NLP

- Transformers are being adapted to other domains, such as vision.
- Positionally-encoded image pieces are passed into an attention layer.
- Analogously to language-modeling, vision transformers can be pre-trained by predicting masked image patches.
- Multi-modal models are being trained, with text and images together.
- Transformers are also behind AlphaFold, breakthrough protein-folding model.




COVID-19

Science

NEWS | BIOLOGY

'The game has changed.' AI triumphs at solving protein structures

In milestone, software predictions finally match structures calculated from experimental data

30 NOV 2020 · BY ROBERT F. SERVICE

Looking back at STAT 4710

Themes of the class, and a lingering question

1. Regression and classification
2. Training predictive models
3. Model complexity
4. Bias-variance trade-off
5. Model selection and model assessment
6. Interpretability of predictive models
7. R programming tools
8. Working with data

Conceptual

Practical

Themes of the class, and a lingering question

1. Regression and classification
2. Training predictive models
3. Model complexity
4. Bias-variance trade-off
5. Model selection and model assessment
6. Interpretability of predictive models
7. R programming tools
8. Working with data

Conceptual

Practical

Lingering question: What is the best prediction method?

Theme: Regression and classification

Theme: Regression and classification

Prediction methods vary based on the response type:

- Regression: continuous responses
- Classification: discrete responses (binary or multi-class)

Theme: Regression and classification

Prediction methods vary based on the response type:

- Regression: continuous responses
- Classification: discrete responses (binary or multi-class)

Most methods have versions for regression and classification, e.g. linear regression and logistic regression.

Theme: Regression and classification

Prediction methods vary based on the response type:

- Regression: continuous responses
- Classification: discrete responses (binary or multi-class)

Most methods have versions for regression and classification, e.g. linear regression and logistic regression.

Classification methods are indirect in the sense that they predict probabilities of each class. They are also a little more fussy; need to make sure probabilities are between 0 and 1, class imbalance, misclassification error versus Gini index,…

Theme: Regression and classification

Prediction methods vary based on the response type:

- Regression: continuous responses
- Classification: discrete responses (binary or multi-class)

Most methods have versions for regression and classification, e.g. linear regression and logistic regression.

Classification methods are indirect in the sense that they predict probabilities of each class. They are also a little more fussy; need to make sure probabilities are between 0 and 1, class imbalance, misclassification error versus Gini index,…

Many of the same intuitions apply for regression and classification.

Theme: Training predictive models

Define class of predictive models $f_\beta(X)$ indexed by some parameter vector β.

Find member of this class that best fits the training data, as measured by the loss function L of predictions given true responses, possibly regularized:

$$\widehat{\beta} = \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} L(Y_i, f_\beta(X_i)) + \lambda \cdot \text{penalty}(\beta).$$

Theme: Training predictive models

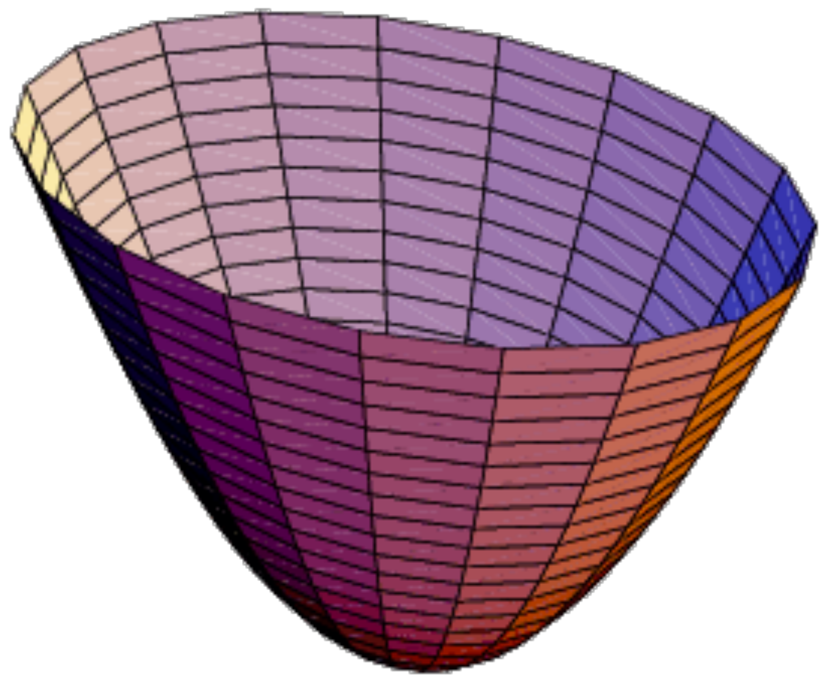
Define class of predictive models $f_\beta(X)$ indexed by some parameter vector β.

Find member of this class that best fits the training data, as measured by the loss function L of predictions given true responses, possibly regularized:

$$\hat{\beta} = \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} L(Y_i, f_\beta(X_i)) + \lambda \cdot \text{penalty}(\beta).$$

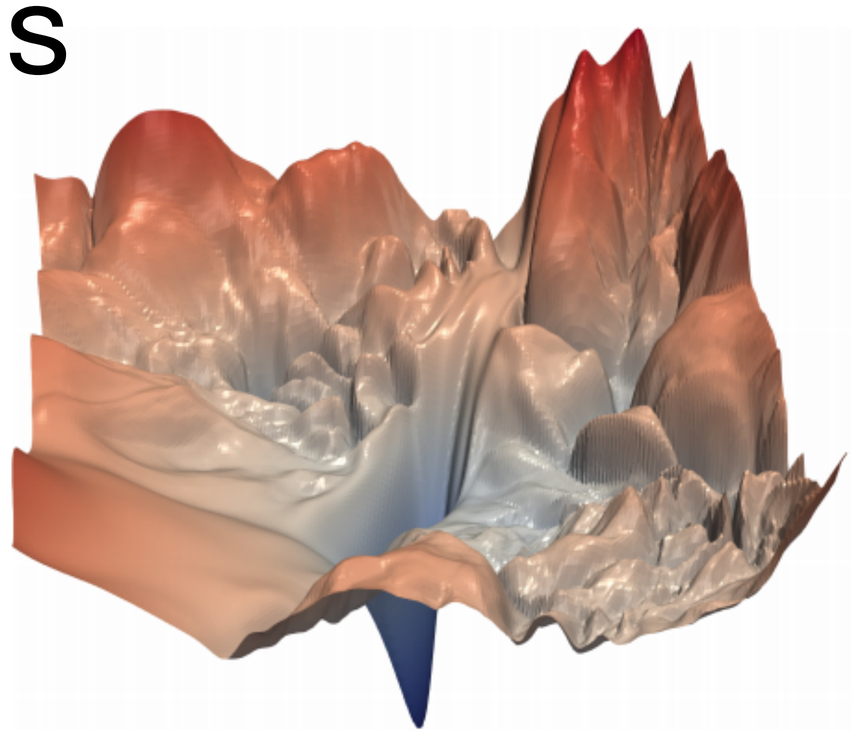
Convex (optimization is easy)

- Linear and logistic regression
- Linear and logistic regression with ridge or lasso penalties



Not convex (optimization is hard)

- Tree-based methods
- Neural networks



https://arxiv.org/abs/1712.09913

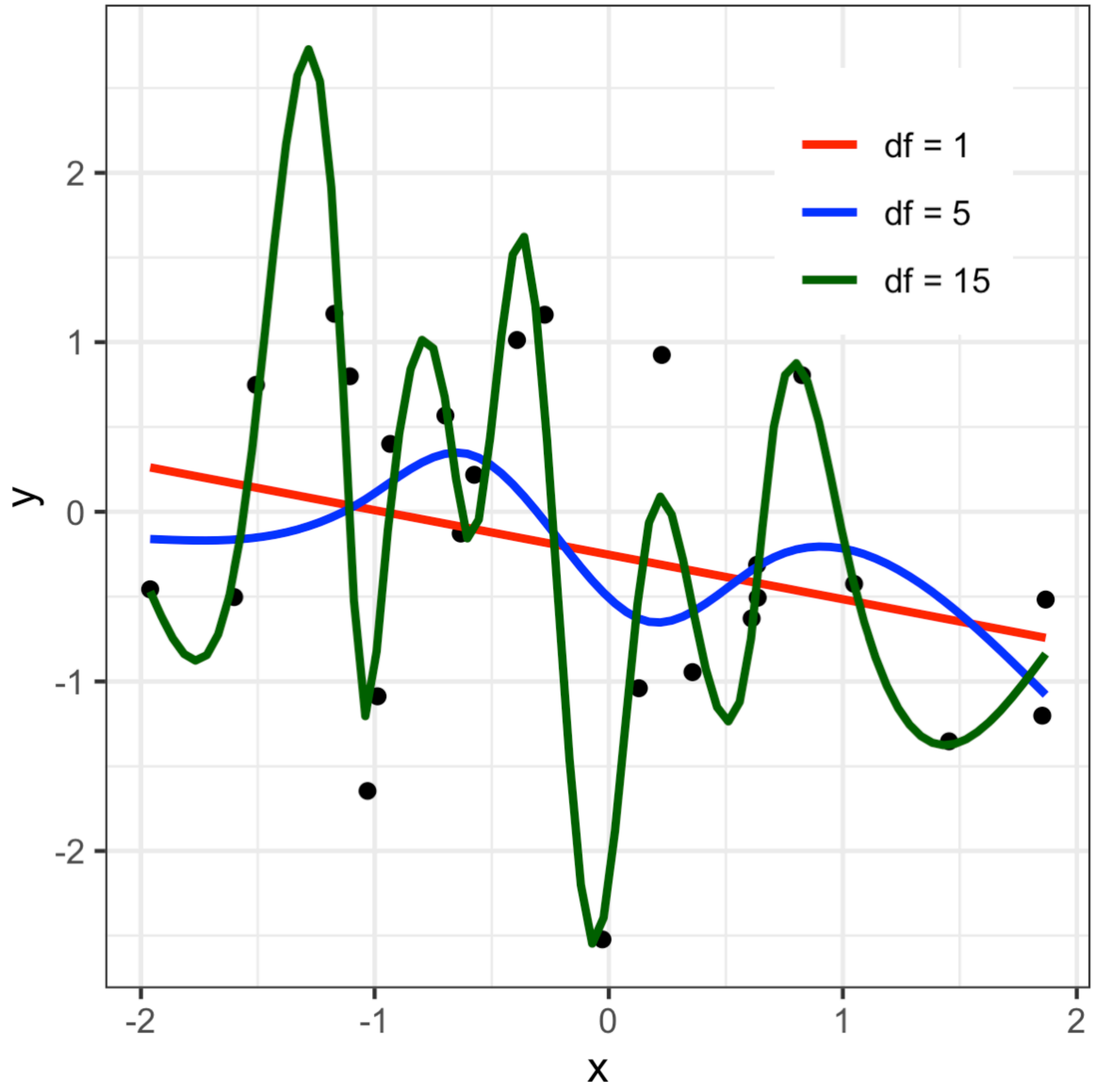
Theme: Model complexity

Theme: Model complexity

One of the key characteristics of a predictive model is its complexity: how flexibly does it fit the training data?

Theme: Model complexity

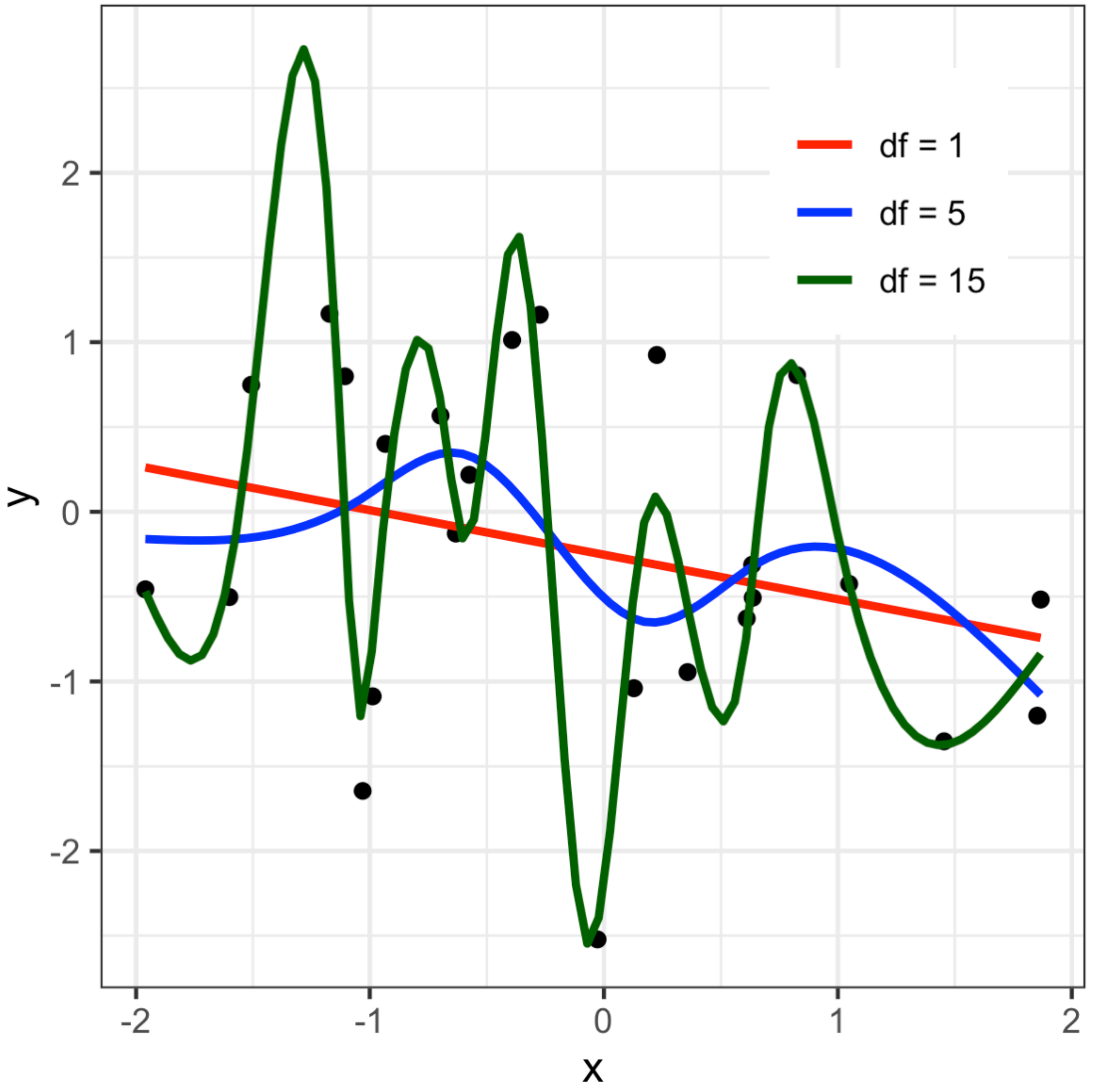
One of the key characteristics of a predictive model is its complexity: how flexibly does it fit the training data?



Theme: Model complexity

One of the key characteristics of a predictive model is its complexity: how flexibly does it fit the training data?

How is model complexity defined? Depends on model:

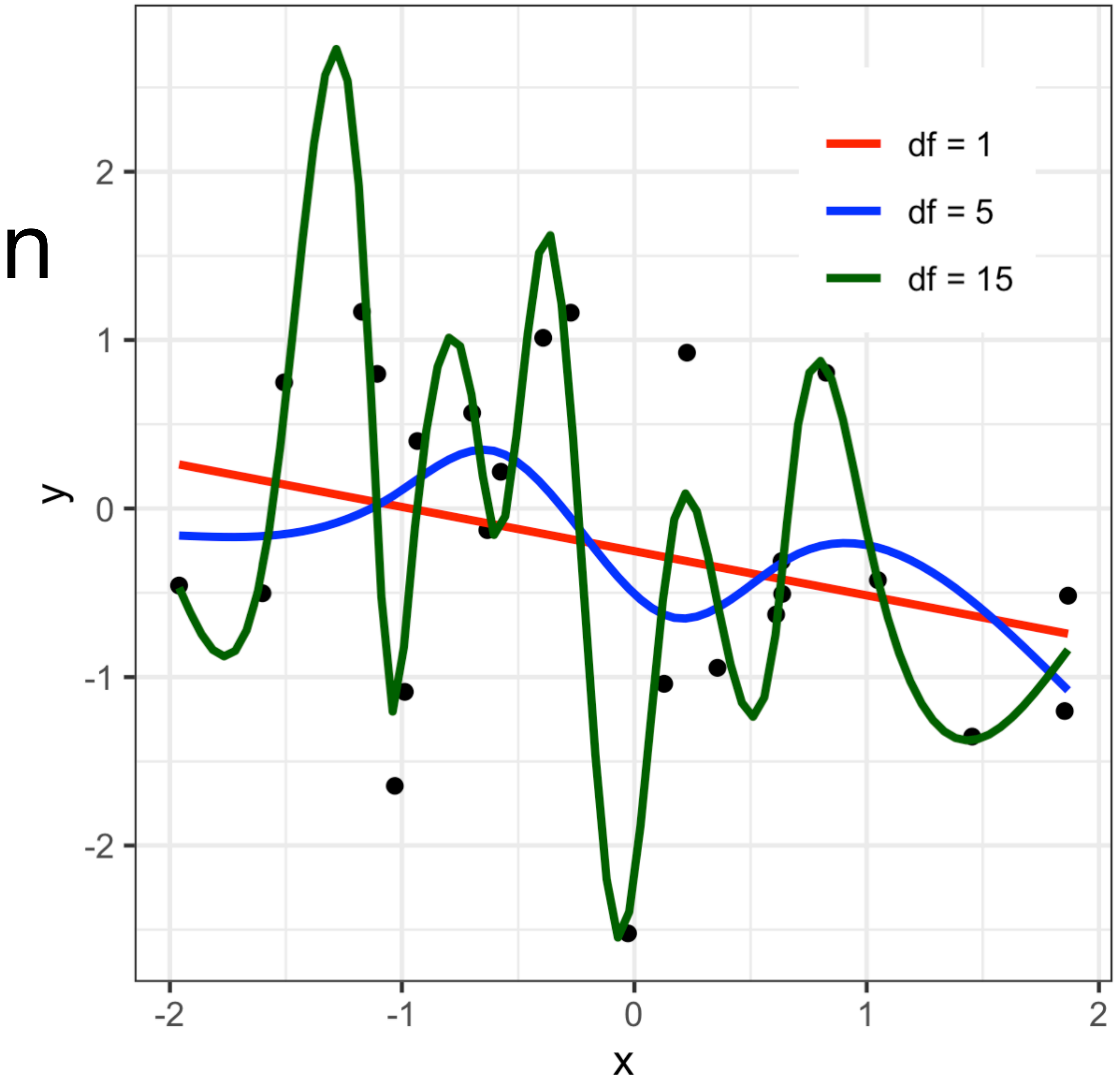


Theme: Model complexity

One of the key characteristics of a predictive model is its complexity: how flexibly does it fit the training data?

How is model complexity defined? Depends on model:

- Number of parameters in linear or logistic regression

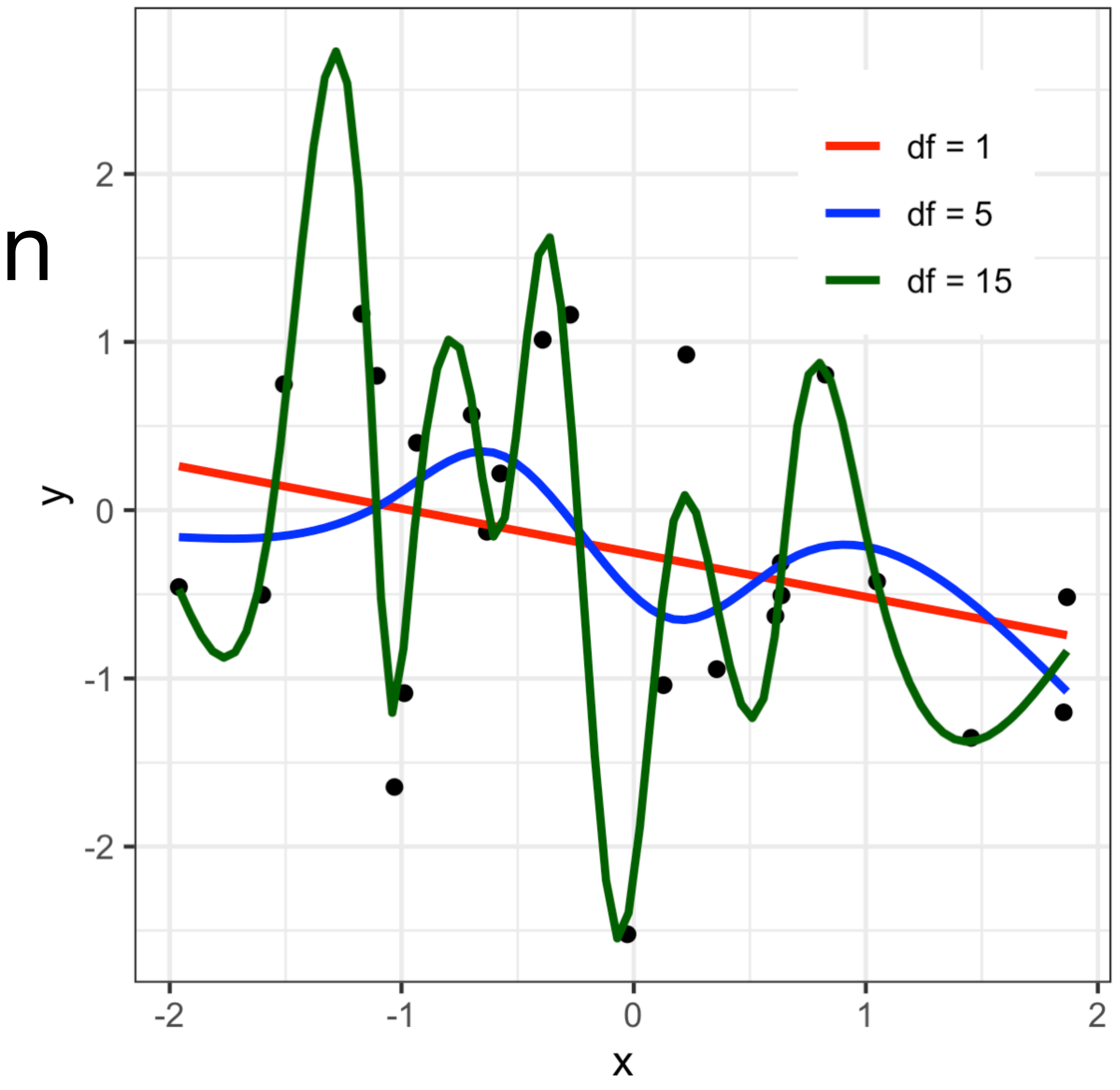


Theme: Model complexity

One of the key characteristics of a predictive model is its complexity: how flexibly does it fit the training data?

How is model complexity defined? Depends on model:

- Number of parameters in linear or logistic regression
- Depth of a decision tree

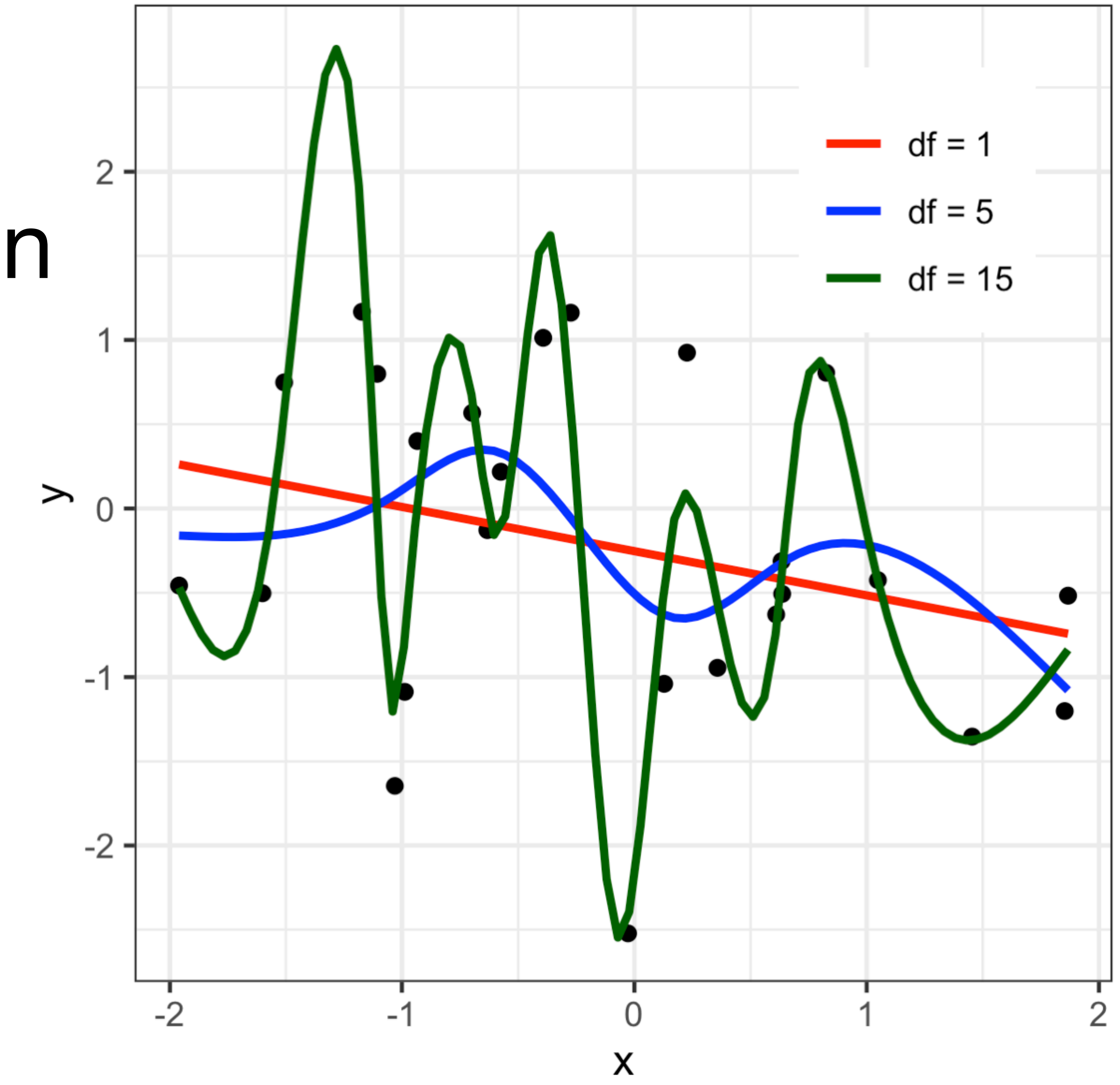


Theme: Model complexity

One of the key characteristics of a predictive model is its complexity: how flexibly does it fit the training data?

How is model complexity defined? Depends on model:

- Number of parameters in linear or logistic regression
- Depth of a decision tree
- Number of neighbors used in K-nearest-neighbors



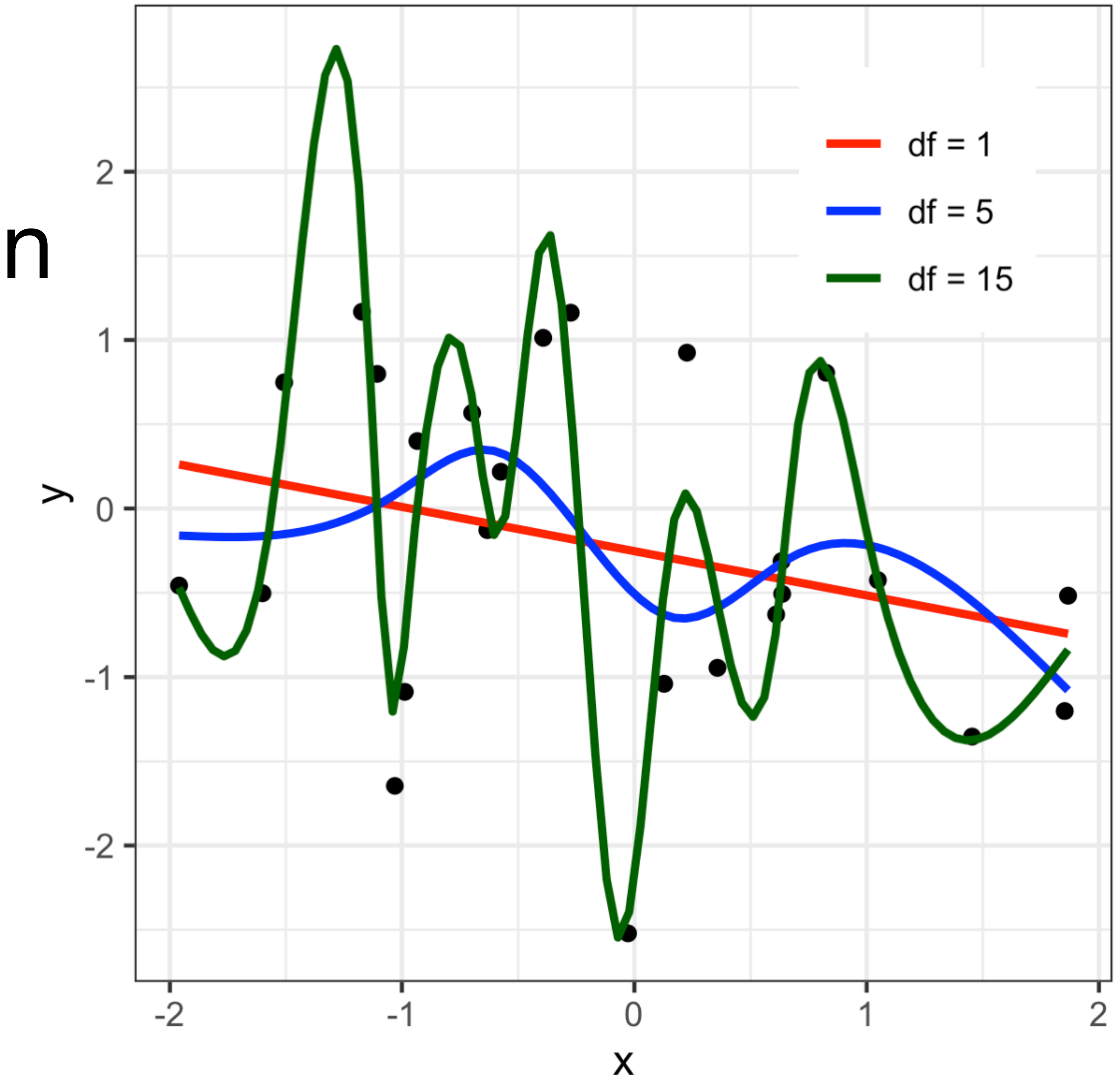
Theme: Model complexity

One of the key characteristics of a predictive model is its complexity: how flexibly does it fit the training data?

How is model complexity defined? Depends on model:

- Number of parameters in linear or logistic regression
- Depth of a decision tree
- Number of neighbors used in K-nearest-neighbors

How is model complexity controlled (regularization)?



Theme: Model complexity

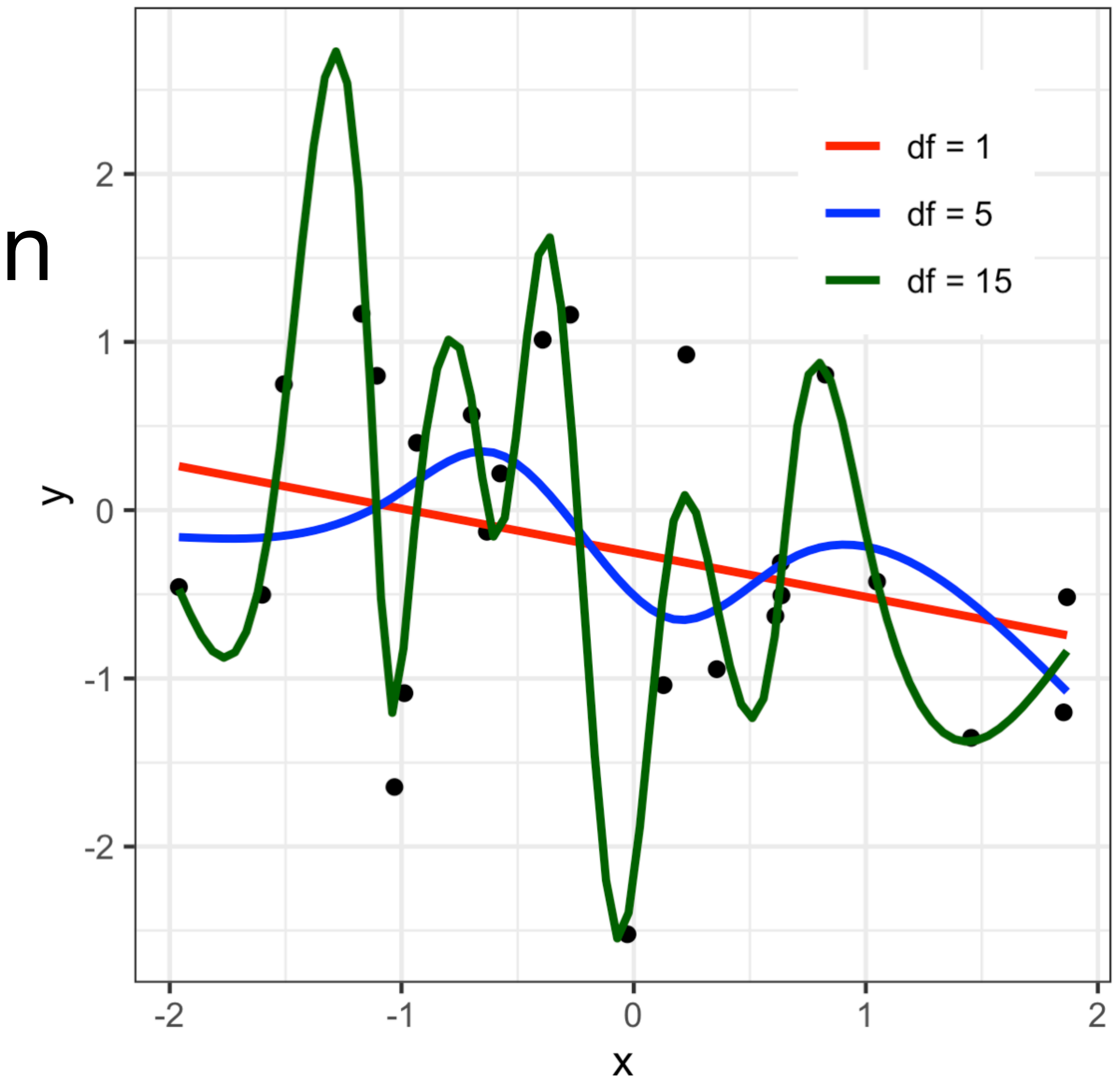
One of the key characteristics of a predictive model is its complexity: how flexibly does it fit the training data?

How is model complexity defined? Depends on model:

- Number of parameters in linear or logistic regression
- Depth of a decision tree
- Number of neighbors used in K-nearest-neighbors

How is model complexity controlled (regularization)?

- Explicit regularization via penalization (lasso, ridge)



Theme: Model complexity

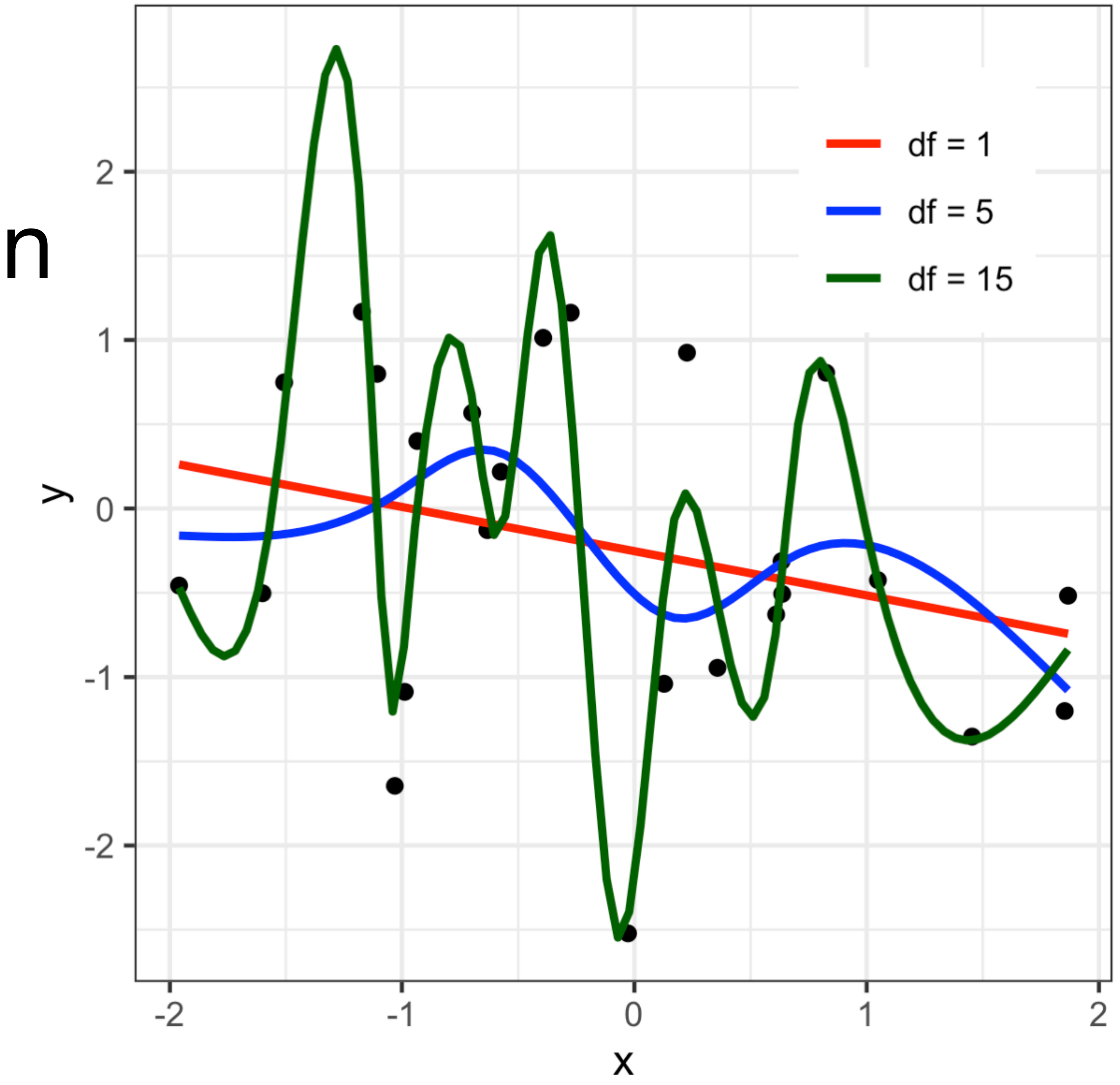
One of the key characteristics of a predictive model is its complexity: how flexibly does it fit the training data?

How is model complexity defined? Depends on model:

- Number of parameters in linear or logistic regression
- Depth of a decision tree
- Number of neighbors used in K-nearest-neighbors

How is model complexity controlled (regularization)?

- Explicit regularization via penalization (lasso, ridge)
- Implicit regularization, e.g. sub-sampling features during random forest model training

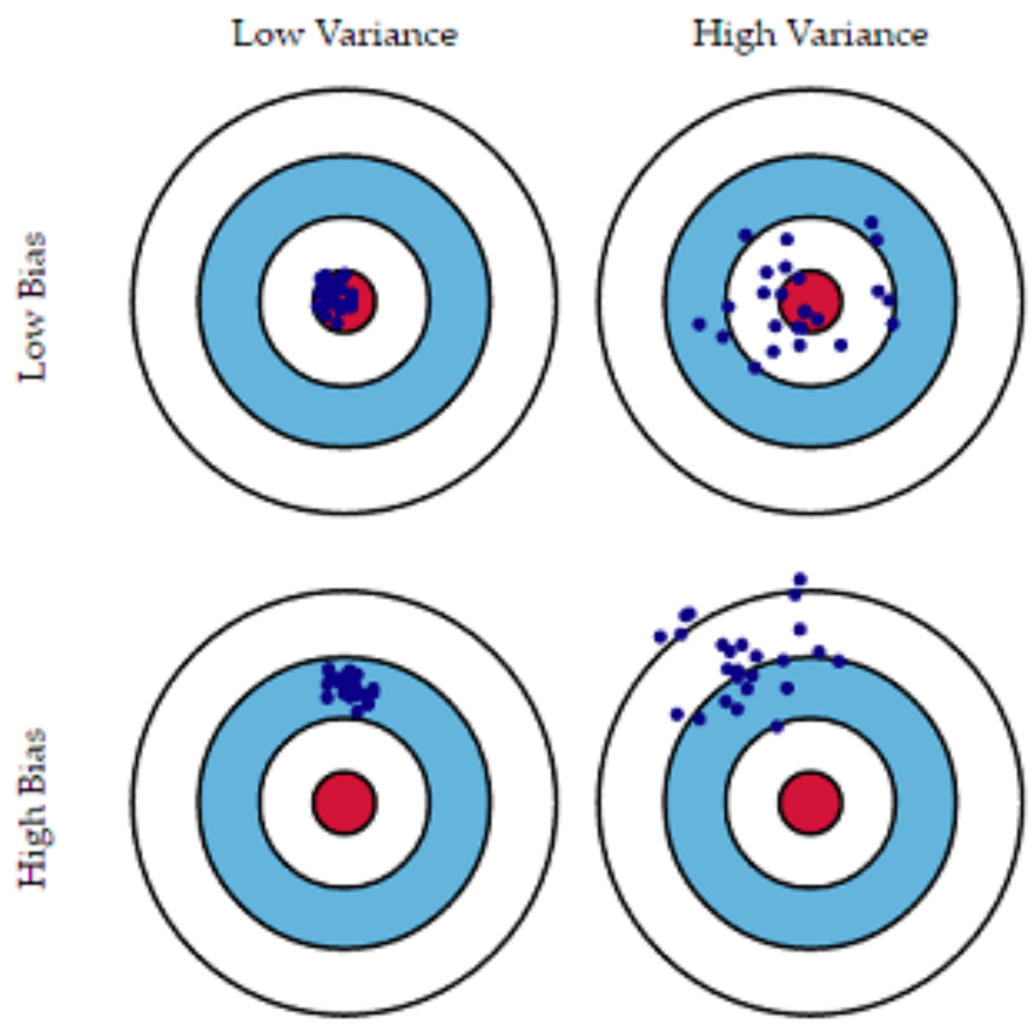


Theme: Bias-variance trade-off

Theme: Bias-variance trade-off

Consider sampling many different training sets.

- Bias: How far off are predictions on average?
- Variance: How much do the predictions wobble around?



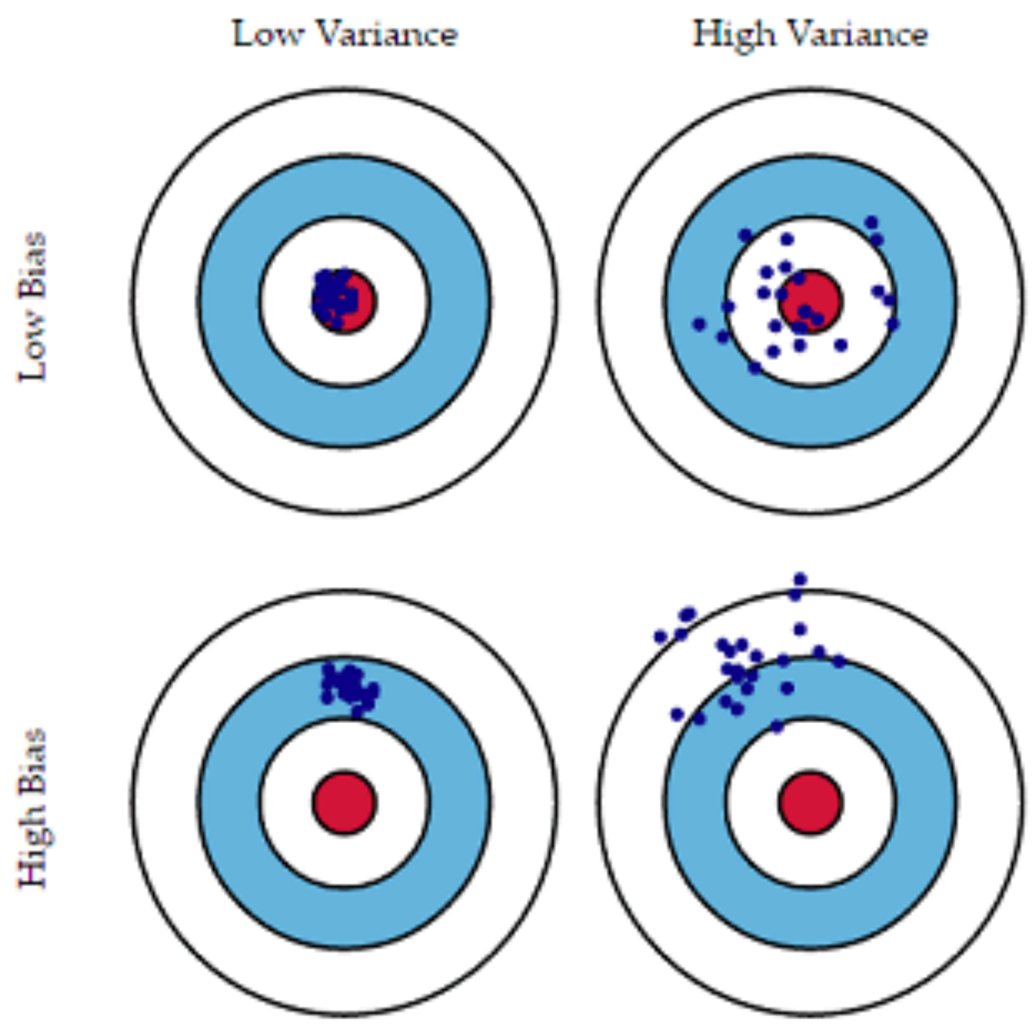
https://www.listendata.com/2017/02/bias-variance-tradeoff.html

Theme: Bias-variance trade-off

Consider sampling many different training sets.

- Bias: How far off are predictions on average?
- Variance: How much do the predictions wobble around?

Higher model complexity leads to less bias but more variance.



https://www.listendata.com/2017/02/bias-variance-tradeoff.html

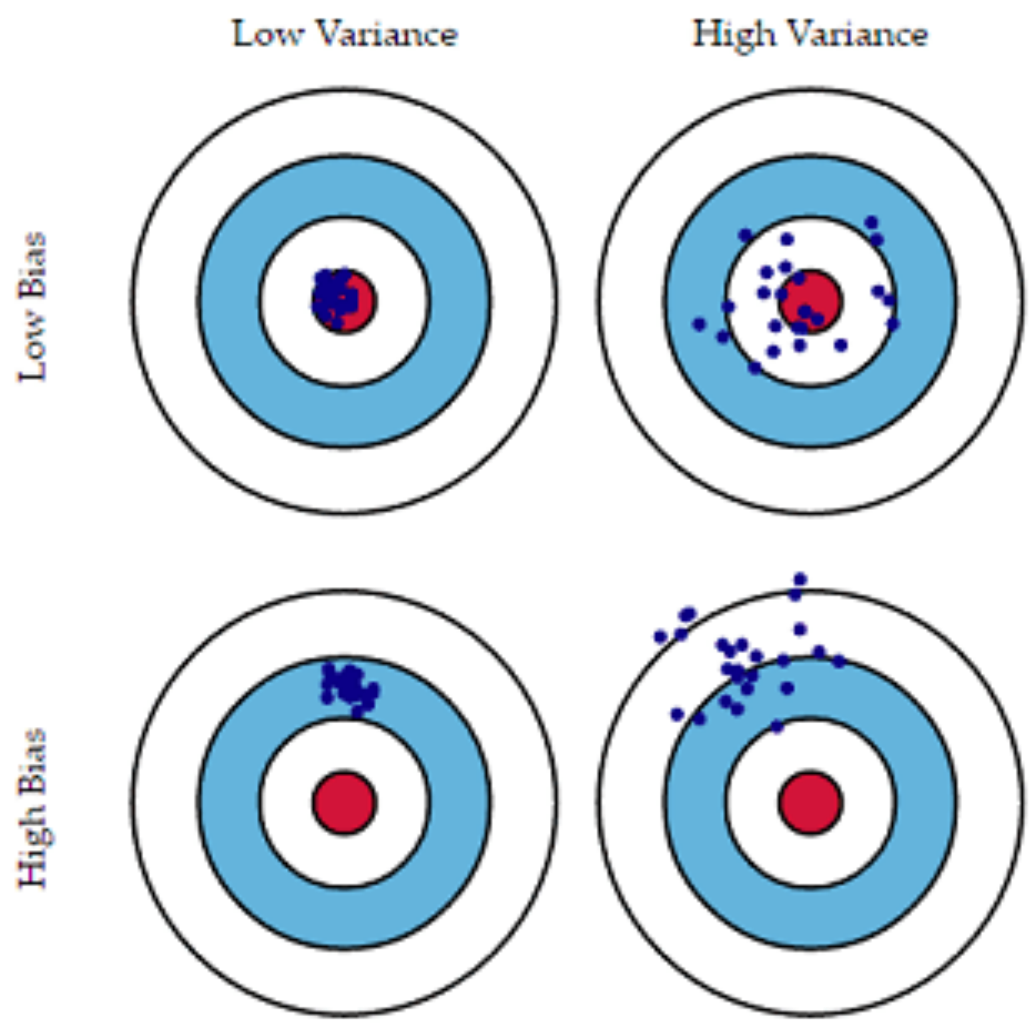
Theme: Bias-variance trade-off

Consider sampling many different training sets.

- Bias: How far off are predictions on average?
- Variance: How much do the predictions wobble around?

Higher model complexity leads to less bias but more variance.

Prediction error = Bias^2+Variance.



https://www.listendata.com/2017/02/bias-variance-tradeoff.html

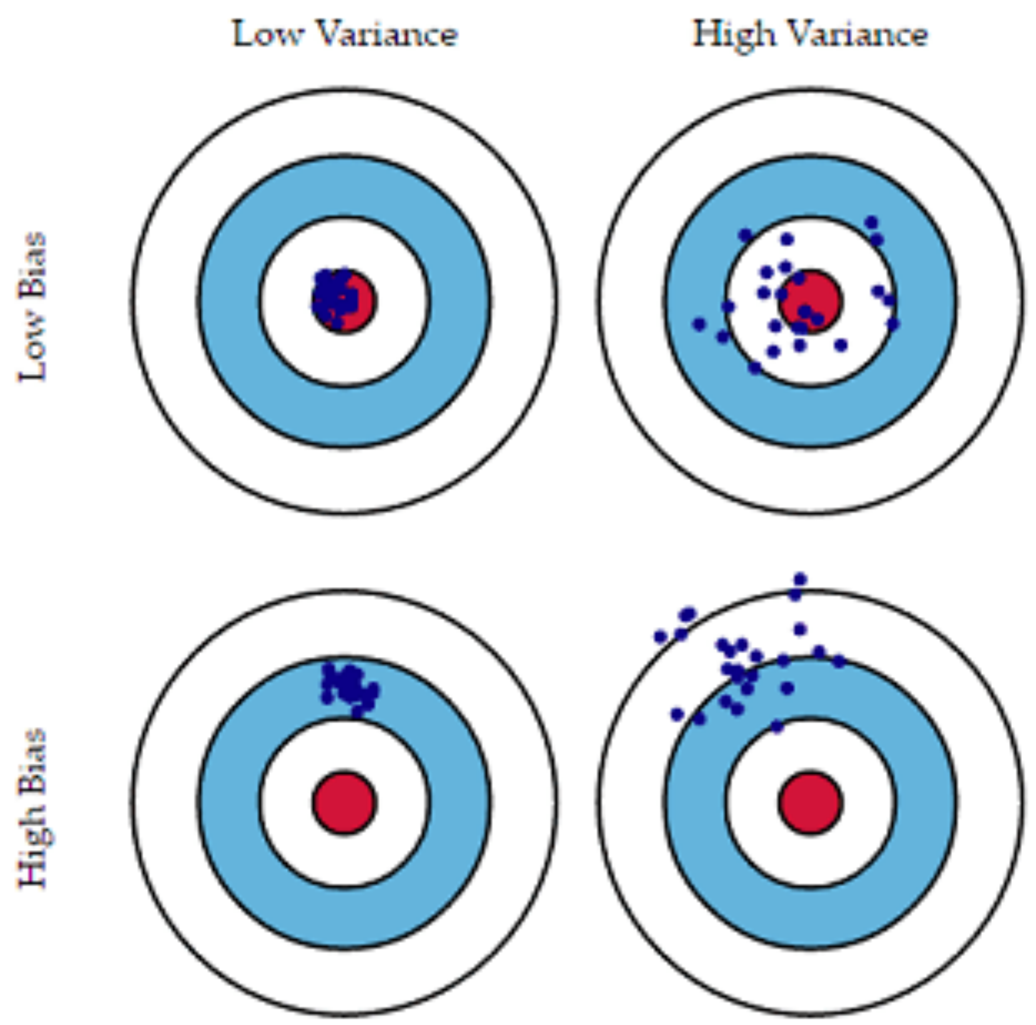
Theme: Bias-variance trade-off

Consider sampling many different training sets.

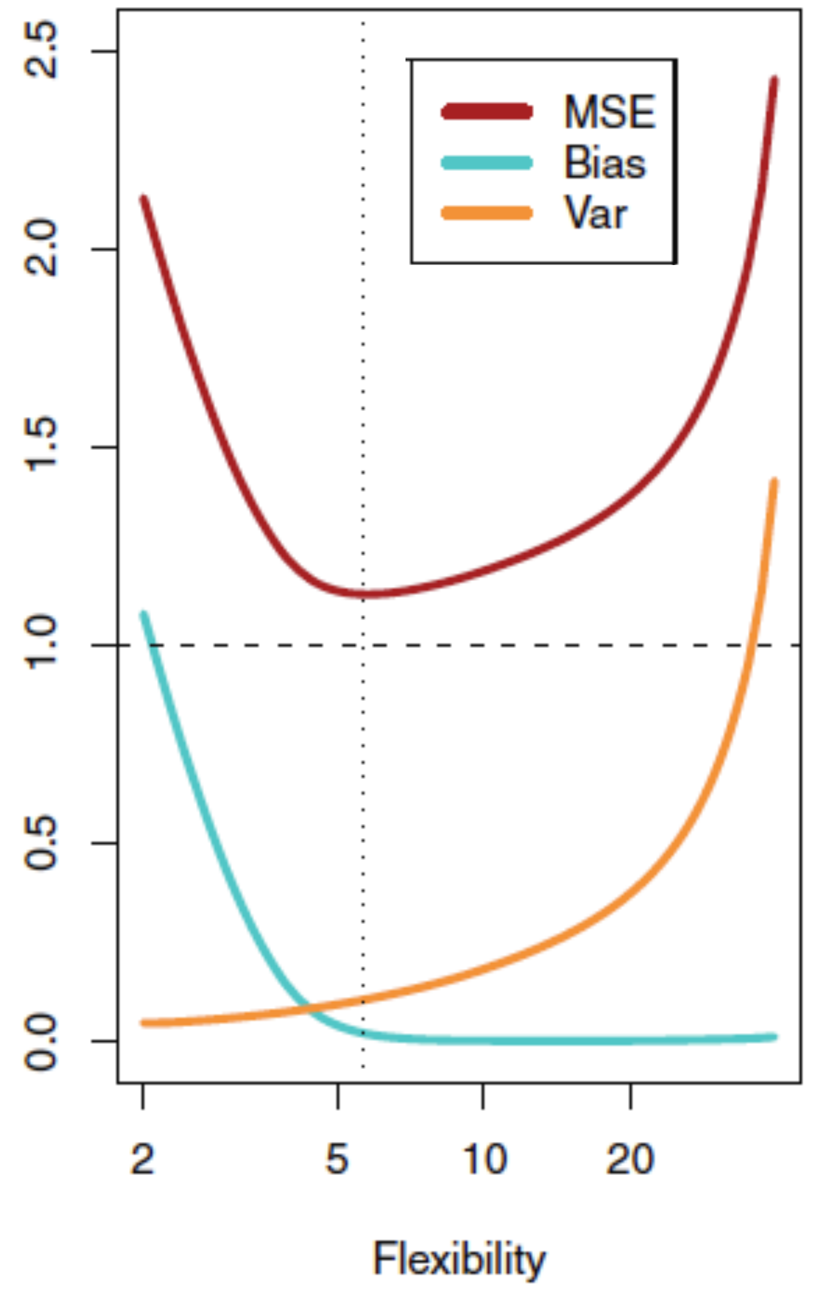
- Bias: How far off are predictions on average?
- Variance: How much do the predictions wobble around?

Higher model complexity leads to less bias but more variance.

Prediction error = Bias^2+Variance.



https://www.listendata.com/2017/02/bias-variance-tradeoff.html



Theme: Bias-variance trade-off

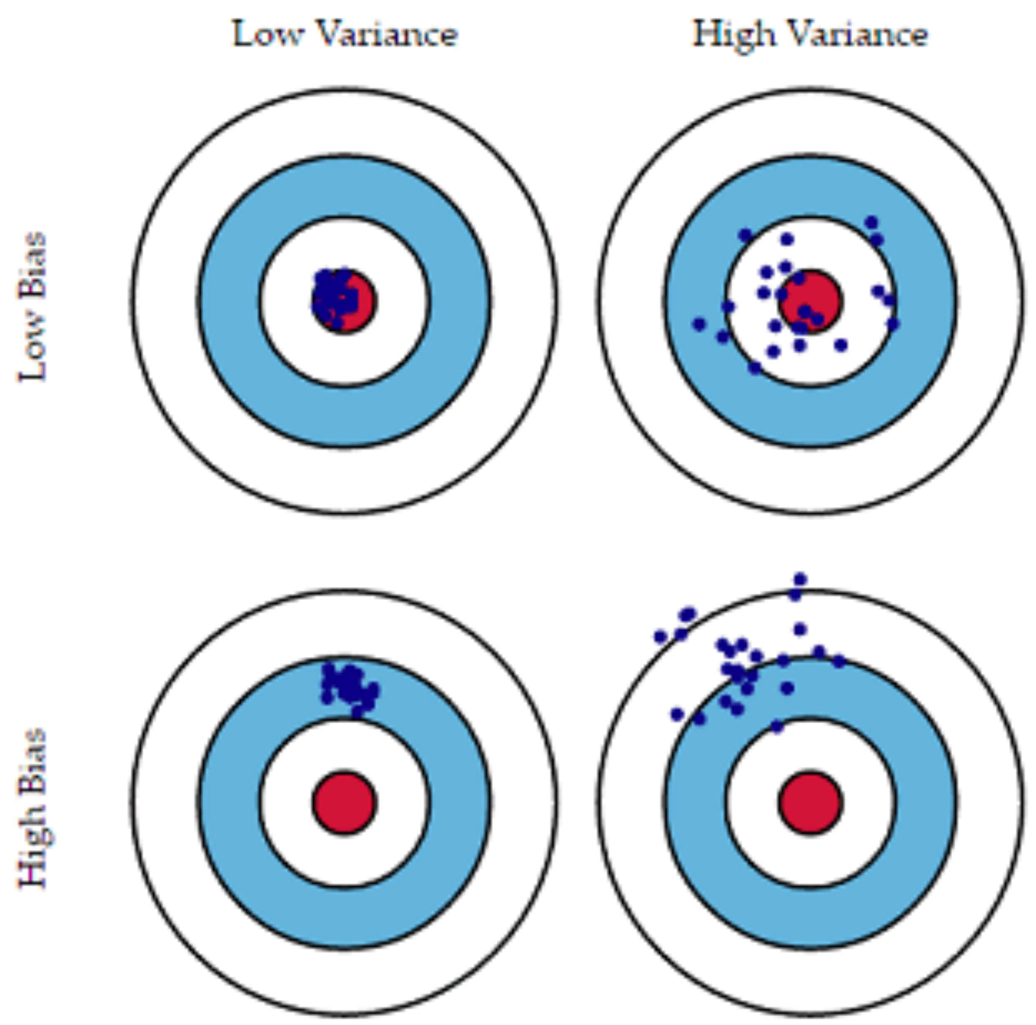
Consider sampling many different training sets.

- Bias: How far off are predictions on average?
- Variance: How much do the predictions wobble around?

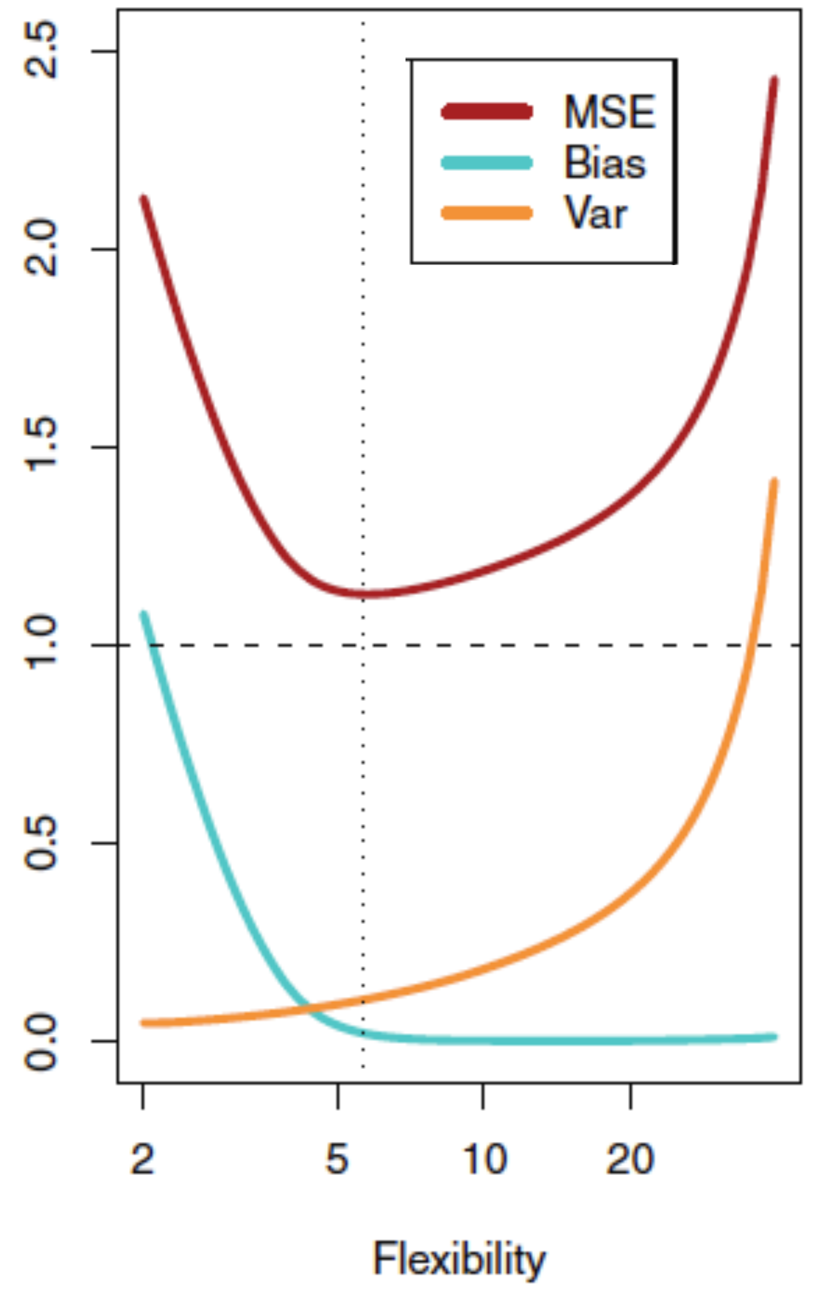
Higher model complexity leads to less bias but more variance.

Prediction error = Bias^2+Variance.

Overfitting: complex models' sensitivity to noise in the training data (high variance) → low training error but high test error.



https://www.listendata.com/2017/02/bias-variance-tradeoff.html



Theme: Bias-variance trade-off

Consider sampling many different training sets.

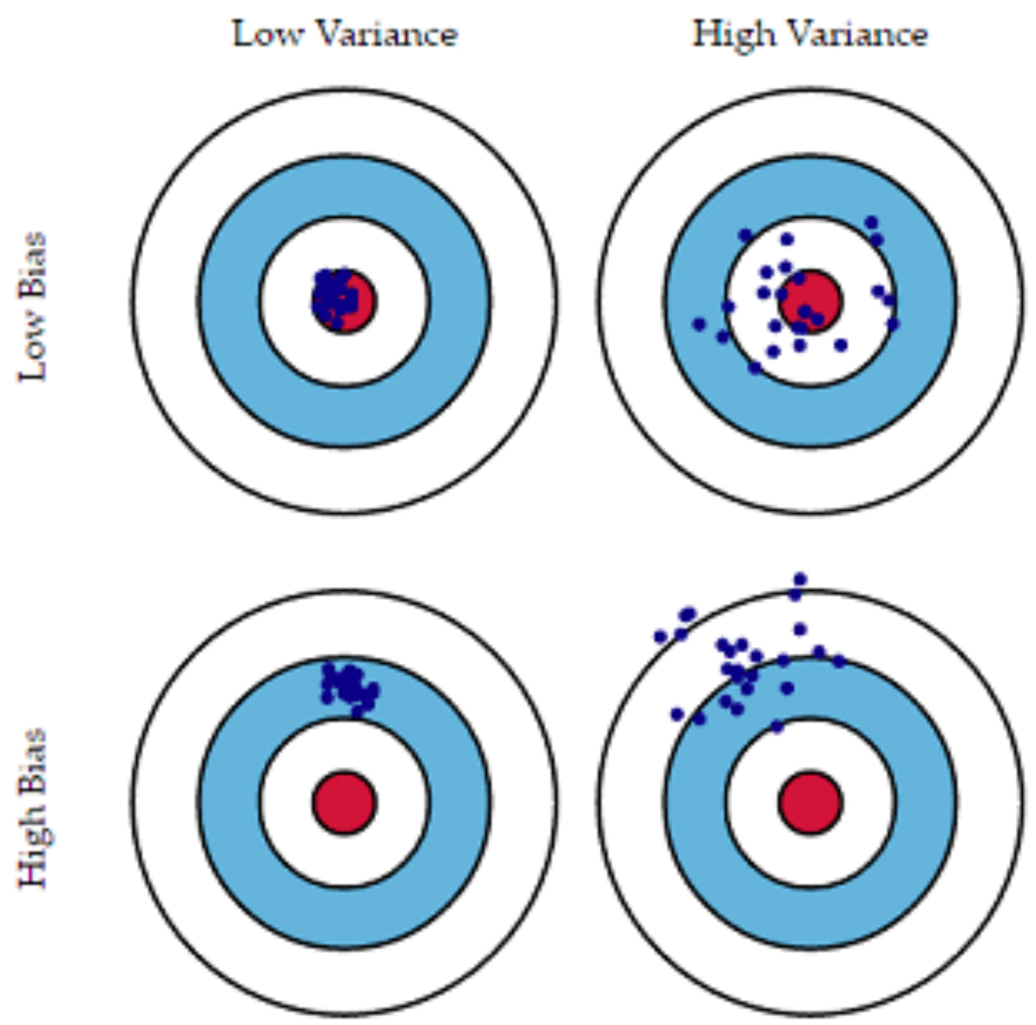
- Bias: How far off are predictions on average?
- Variance: How much do the predictions wobble around?

Higher model complexity leads to less bias but more variance.

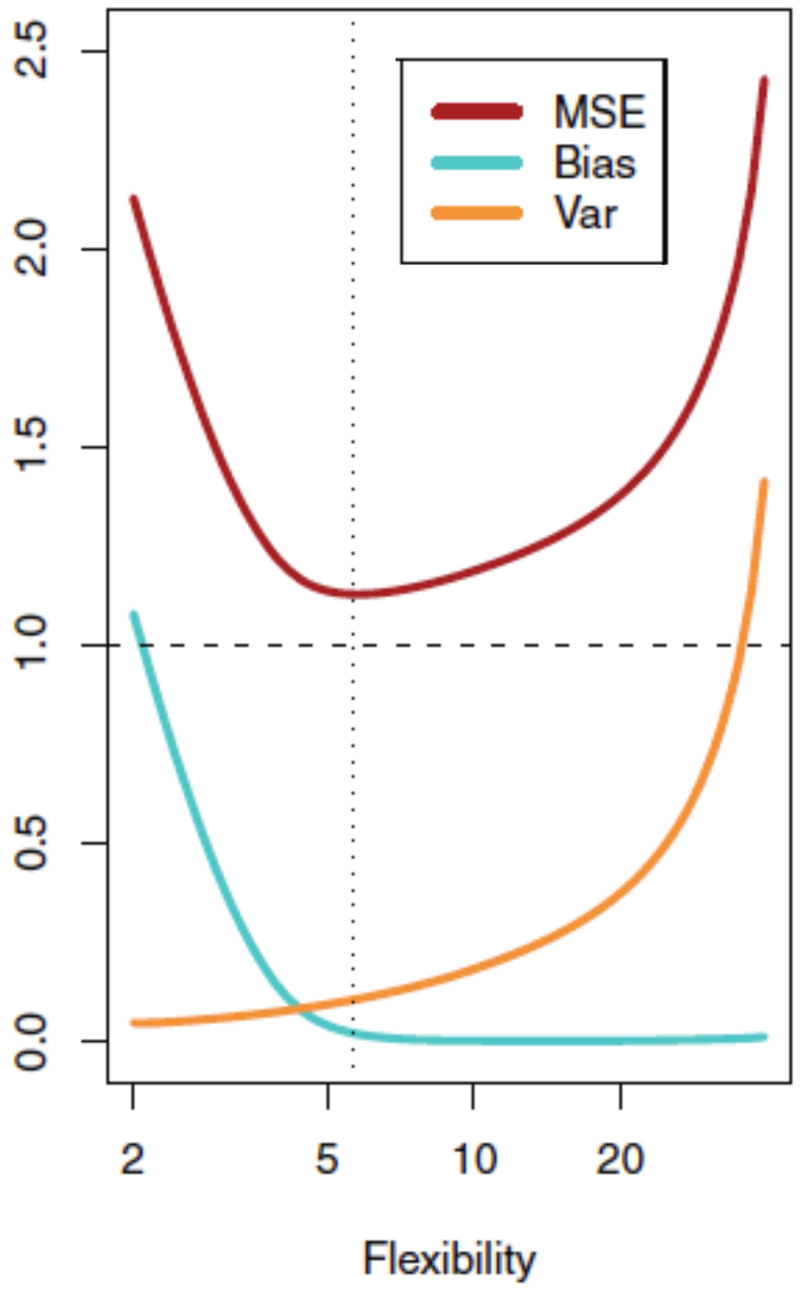
Prediction error = $\text{Bias}^2 + \text{Variance}$.

Overfitting: complex models' sensitivity to noise in the training data (high variance) → low training error but high test error.

Variance increases as noise variance increases, model flexibility increases, training sample size decreases.



https://www.listendata.com/2017/02/bias-variance-tradeoff.html



Theme: Bias-variance trade-off

Consider sampling many different training sets.

- Bias: How far off are predictions on average?
- Variance: How much do the predictions wobble around?

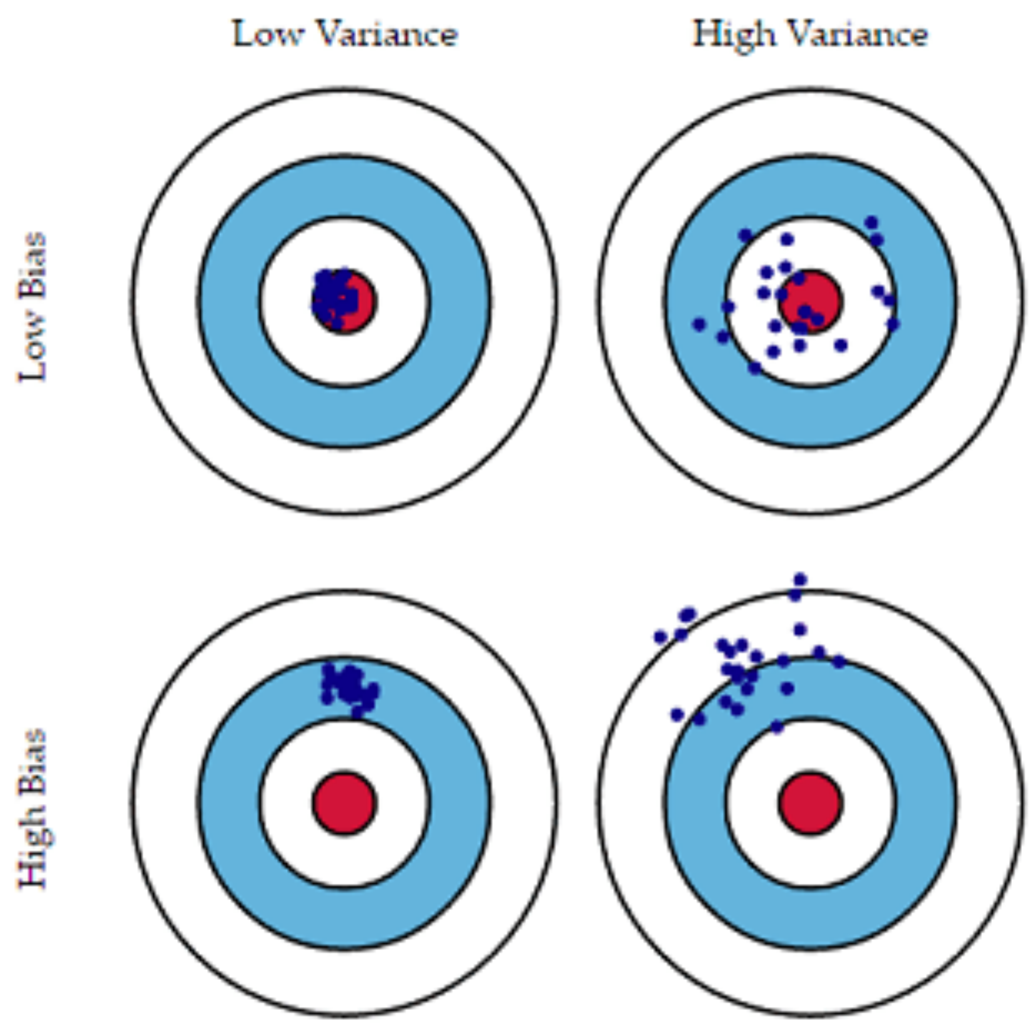
Higher model complexity leads to less bias but more variance.

Prediction error = $\text{Bias}^2 + \text{Variance}$.

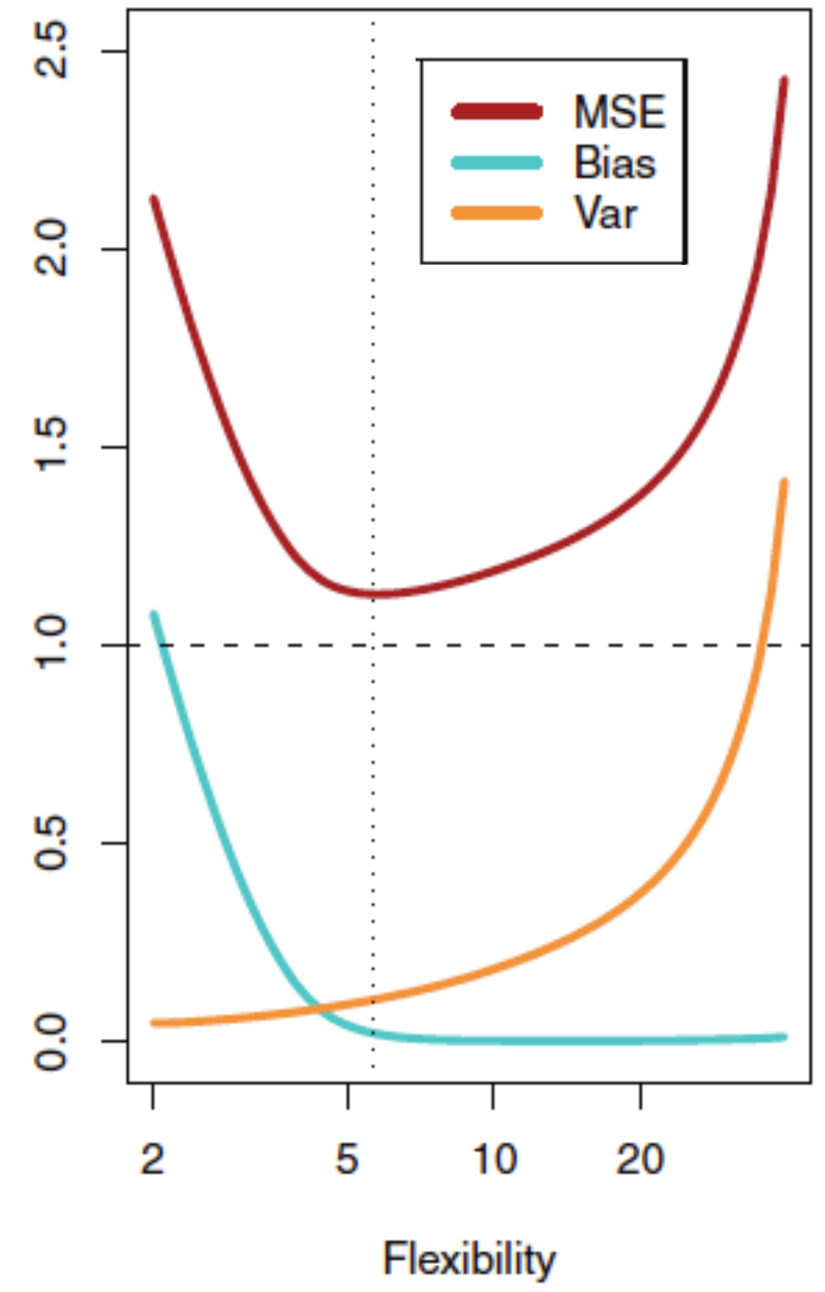
Overfitting: complex models' sensitivity to noise in the training data (high variance) → low training error but high test error.

Variance increases as noise variance increases, model flexibility increases, training sample size decreases.

Lower noise or larger sample size means you can afford more complex model (think of deep learning).



https://www.listendata.com/2017/02/bias-variance-tradeoff.html



Theme: Model selection and model assessment

Theme: Model selection and model assessment

Model assessment

Theme: Model selection and model assessment

Model assessment

- Predictive models are assessed based on test data that is separate from the data used to train these models.

Theme: Model selection and model assessment

Model assessment

- Predictive models are assessed based on test data that is separate from the data used to train these models.
- Different criteria are used to quantify the accuracy of predictions, like RMSE, misclassification error, F-score, and confusion matrix.

Theme: Model selection and model assessment

Model assessment

- Predictive models are assessed based on test data that is separate from the data used to train these models.
- Different criteria are used to quantify the accuracy of predictions, like RMSE, misclassification error, F-score, and confusion matrix.
- Model assessment can be subtle for classification problems.

Theme: Model selection and model assessment

Model assessment

- Predictive models are assessed based on test data that is separate from the data used to train these models.
- Different criteria are used to quantify the accuracy of predictions, like RMSE, misclassification error, F-score, and confusion matrix.
- Model assessment can be subtle for classification problems.

Model selection

Theme: Model selection and model assessment

Model assessment

- Predictive models are assessed based on test data that is separate from the data used to train these models.
- Different criteria are used to quantify the accuracy of predictions, like RMSE, misclassification error, F-score, and confusion matrix.
- Model assessment can be subtle for classification problems.

Model selection

- Main tool: Cross-validation, which mimics the train-test split using folds.

Theme: Model selection and model assessment

Model assessment

- Predictive models are assessed based on test data that is separate from the data used to train these models.
- Different criteria are used to quantify the accuracy of predictions, like RMSE, misclassification error, F-score, and confusion matrix.
- Model assessment can be subtle for classification problems.

Model selection

- Main tool: Cross-validation, which mimics the train-test split using folds.
- Other schemes for model assessment: validation set approach and out-of-bag error, the latter for random forests.

Theme: Model selection and model assessment

Model assessment

- Predictive models are assessed based on test data that is separate from the data used to train these models.
- Different criteria are used to quantify the accuracy of predictions, like RMSE, misclassification error, F-score, and confusion matrix.
- Model assessment can be subtle for classification problems.

Model selection

- Main tool: Cross-validation, which mimics the train-test split using folds.
- Other schemes for model assessment: validation set approach and out-of-bag error, the latter for random forests.
- One-standard-error rule reflects preference for simpler models.

Theme: Interpretability of predictive models

We want to understand how our predictive model is arriving at its conclusions.

Theme: Interpretability of predictive models

We want to understand how our predictive model is arriving at its conclusions.

More interpretable Less interpretable

Theme: Interpretability of predictive models

We want to understand how our predictive model is arriving at its conclusions.

More interpretable ⟷ Less interpretable

Can understand the model
by directly inspecting it.

Theme: Interpretability of predictive models

We want to understand how our predictive model is arriving at its conclusions.

Decision
trees

More interpretable

Less interpretable

Can understand the model
by directly inspecting it.

Theme: Interpretability of predictive models

We want to understand how our predictive model is arriving at its conclusions.

Decision trees

Lasso

More interpretable

Less interpretable

Can understand the model by directly inspecting it.

Theme: Interpretability of predictive models

We want to understand how our predictive model is arriving at its conclusions.

Decision trees

Lasso

Linear, logistic regression

More interpretable ⟷ Less interpretable

Can understand the model by directly inspecting it.

Theme: Interpretability of predictive models

We want to understand how our predictive model is arriving at its conclusions.

Decision trees | Lasso | Linear, logistic regression

More interpretable ⟵⟶ Less interpretable

Can understand the model by directly inspecting it.

Try to understand model by summarizing, e.g. feature importance scores.

Theme: Interpretability of predictive models

We want to understand how our predictive model is arriving at its conclusions.

Decision trees | Lasso | Linear, logistic regression | Boosting

More interpretable ⟷ Less interpretable

Can understand the model by directly inspecting it.

Try to understand model by summarizing, e.g. feature importance scores.

Theme: Interpretability of predictive models

We want to understand how our predictive model is arriving at its conclusions.

Decision trees | Lasso | Linear, logistic regression | Boosting | Random forests

More interpretable ⟷ Less interpretable

Can understand the model by directly inspecting it.

Try to understand model by summarizing, e.g. feature importance scores.

Theme: Interpretability of predictive models

We want to understand how our predictive model is arriving at its conclusions.

Decision trees

Lasso

Linear, logistic regression

Boosting

Random forests

Neural networks

More interpretable

Less interpretable

Can understand the model by directly inspecting it.

Try to understand model by summarizing, e.g. feature importance scores.

Theme: R programming tools

Theme: R programming tools

Tools:

Theme: R programming tools

Tools:

- tidyverse is a nice way to import, clean, transform, and visualize data.

Theme: R programming tools

Tools:

- tidyverse is a nice way to import, clean, transform, and visualize data.
- R is well suited for data science; lots of packages available to analyze data.

Theme: R programming tools

Tools:

- tidyverse is a nice way to import, clean, transform, and visualize data.
- R is well suited for data science; lots of packages available to analyze data.
- R Markdown is a nice way to integrate text, code, and output.

Theme: R programming tools

Tools:

- tidyverse is a nice way to import, clean, transform, and visualize data.
- R is well suited for data science; lots of packages available to analyze data.
- R Markdown is a nice way to integrate text, code, and output.
- RStudio Cloud gives a ready-to-go R programming environment.

Theme: R programming tools

Tools:

- tidyverse is a nice way to import, clean, transform, and visualize data.
- R is well suited for data science; lots of packages available to analyze data.
- R Markdown is a nice way to integrate text, code, and output.
- RStudio Cloud gives a ready-to-go R programming environment.

Lessons learned:

Theme: R programming tools

Tools:

- tidyverse is a nice way to import, clean, transform, and visualize data.
- R is well suited for data science; lots of packages available to analyze data.
- R Markdown is a nice way to integrate text, code, and output.
- RStudio Cloud gives a ready-to-go R programming environment.

Lessons learned:

- Programming takes patience, attention to detail, and lots of Googling.

Theme: R programming tools

Tools:

- tidyverse is a nice way to import, clean, transform, and visualize data.
- R is well suited for data science; lots of packages available to analyze data.
- R Markdown is a nice way to integrate text, code, and output.
- RStudio Cloud gives a ready-to-go R programming environment.

Lessons learned:

- Programming takes patience, attention to detail, and lots of Googling.
- Each R package and each software tool has its own quirks and limitations.

Theme: R programming tools

Tools:

- tidyverse is a nice way to import, clean, transform, and visualize data.
- R is well suited for data science; lots of packages available to analyze data.
- R Markdown is a nice way to integrate text, code, and output.
- RStudio Cloud gives a ready-to-go R programming environment.

Lessons learned:

- Programming takes patience, attention to detail, and lots of Googling.
- Each R package and each software tool has its own quirks and limitations.
- With practice, these programming and software tools can be very powerful.

Theme: Working with data

Theme: Working with data

- Though increasingly abundant, data are still a precious resource, more of which gives better predictions.

Theme: Working with data

- Though increasingly abundant, data are still a precious resource, more of which gives better predictions.
- Especially in the real world, data are messy and require cleaning.

Theme: Working with data

- Though increasingly abundant, data are still a precious resource, more of which gives better predictions.
- Especially in the real world, data are messy and require cleaning.
- Exploratory data analysis and visualization can reveal a lot about a dataset.

Theme: Working with data

- Though increasingly abundant, data are still a precious resource, more of which gives better predictions.
- Especially in the real world, data are messy and require cleaning.
- Exploratory data analysis and visualization can reveal a lot about a dataset.
- The most successful analyses couple statistical intuition and data intuition.

Theme: Working with data

- Though increasingly abundant, data are still a precious resource, more of which gives better predictions.
- Especially in the real world, data are messy and require cleaning.
- Exploratory data analysis and visualization can reveal a lot about a dataset.
- The most successful analyses couple statistical intuition and data intuition.
- Ultimate goal of data science is to create knowledge and/or make decisions; we must make conclusions relevant to the underlying real-world problem.

A lingering question: What is the best method?

A lingering question: What is the best method?

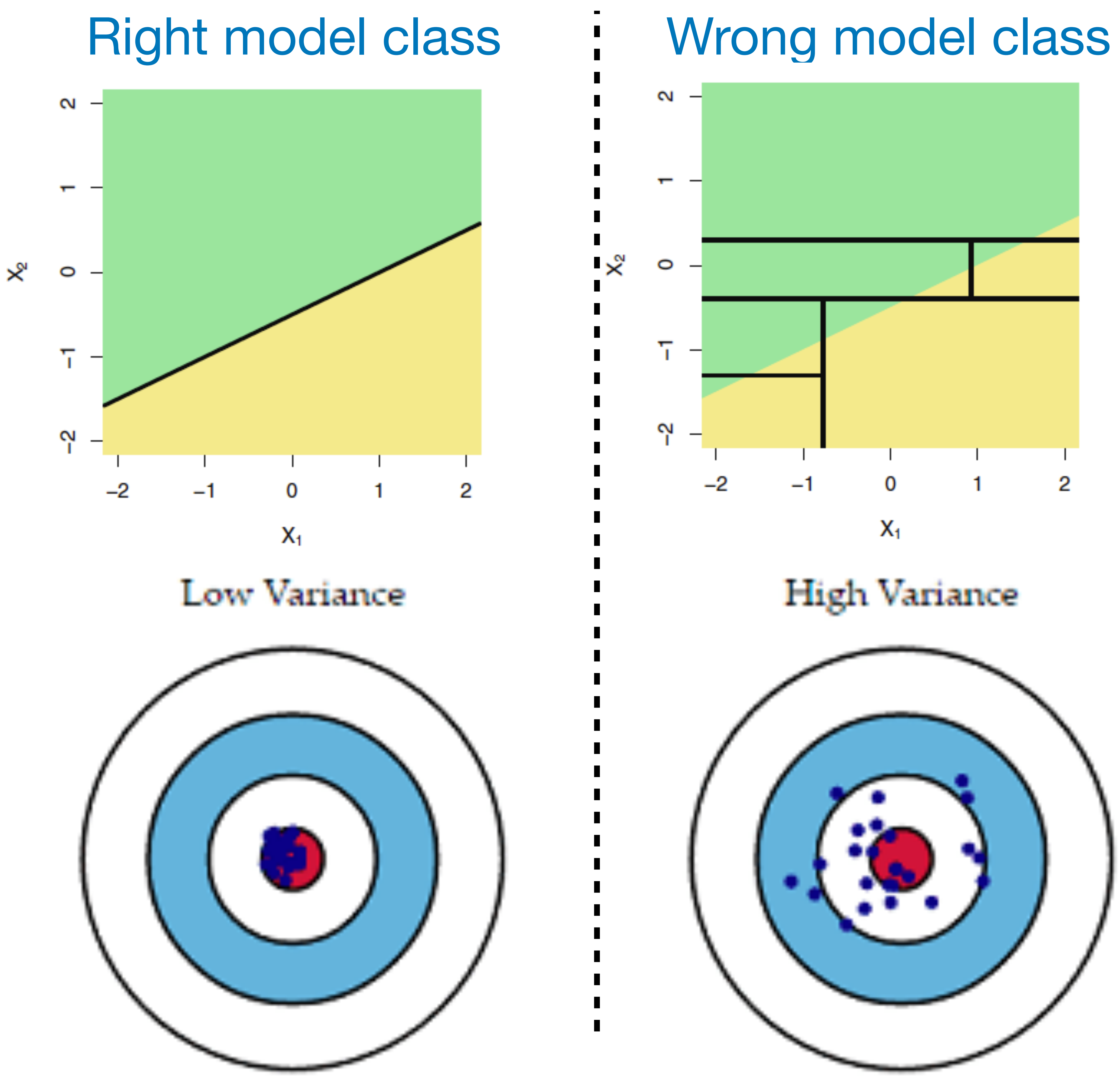
There is no best method. Different methods will work better in different situations.

A lingering question: What is the best method?

There is no best method. Different methods will work better in different situations.

Each prediction method “has in mind” an underlying model class, e.g. linear models for linear regression versus piece-wise constant models for trees.

If true feature-response relationship matches model class our method “has in mind,” will take fewer parameters (less variance) to fit underlying trend (less bias).



https://www.listendata.com/2017/02/bias-variance-tradeoff.html

A lingering question: What is the best method?

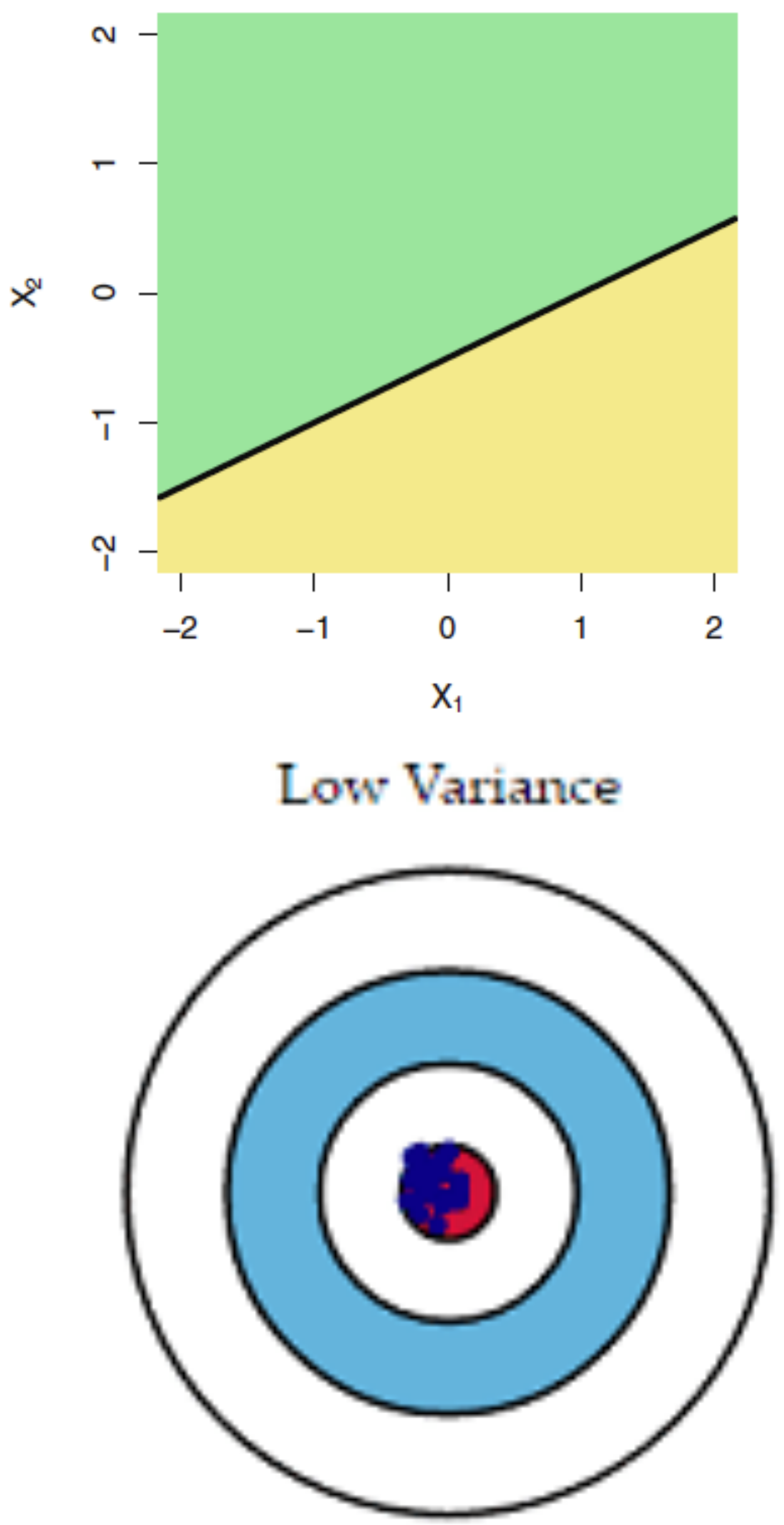
There is no best method. Different methods will work better in different situations.

Each prediction method “has in mind” an underlying model class, e.g. linear models for linear regression versus piece-wise constant models for trees.

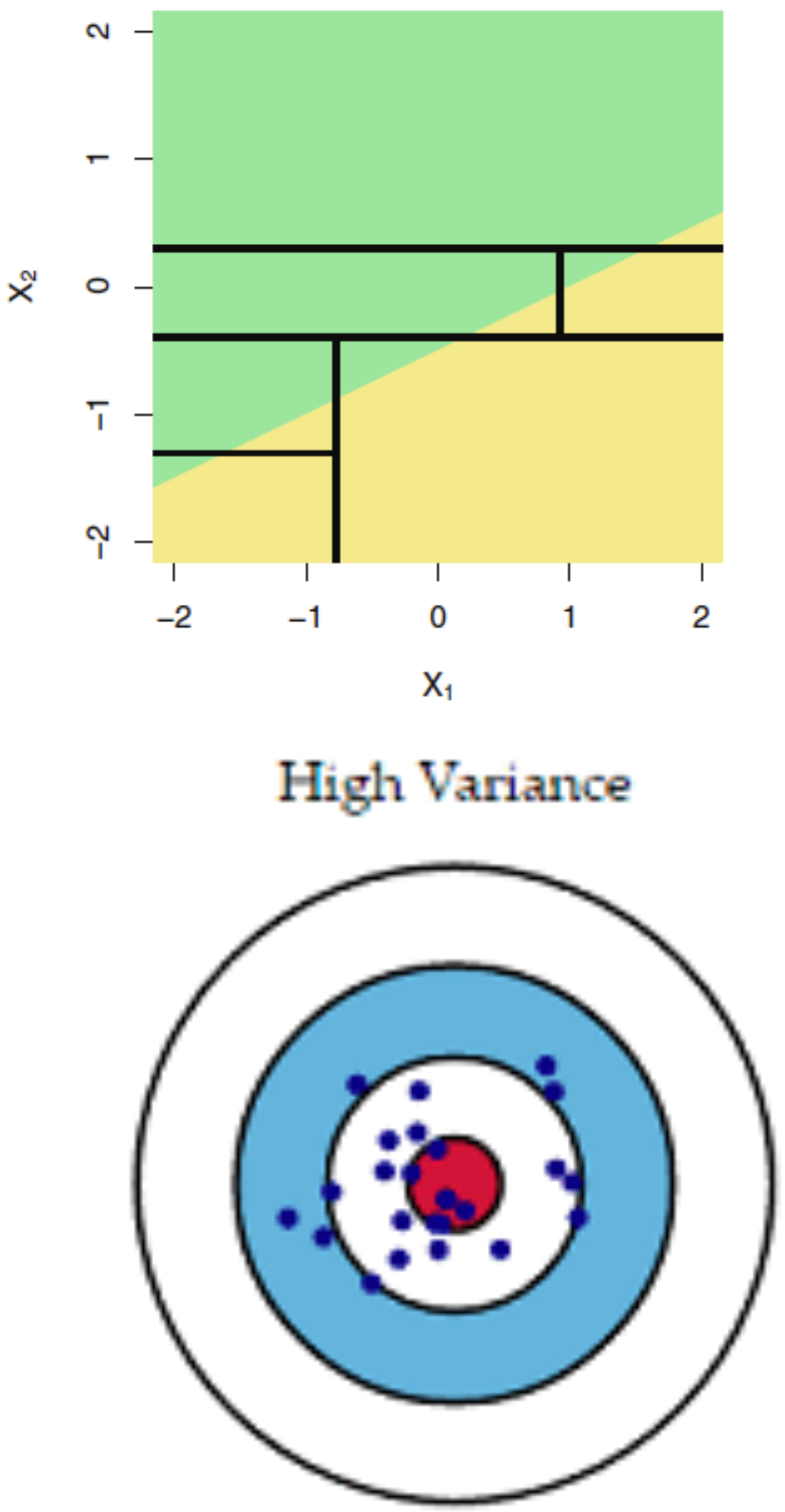
If true feature-response relationship matches model class our method “has in mind,” will take fewer parameters (less variance) to fit underlying trend (less bias).

Moral of the story: It’s good to know several prediction methods. Seek out the ones whose underlying model class you think matches the true feature-response relationship.

Right model class



Wrong model class



https://www.listendata.com/2017/02/bias-variance-tradeoff.html

Looking beyond STAT 4710

What comes next?

What comes next?

- Where are data mining, machine learning, and AI going in the future?

What comes next?

- Where are data mining, machine learning, and AI going in the future?
- How do I learn more about data mining, machine learning, and AI?

What comes next?

- Where are data mining, machine learning, and AI going in the future?
- How do I learn more about data mining, machine learning, and AI?
- What other topics are relevant to data science beyond predictive modeling?

What comes next?

- Where are data mining, machine learning, and AI going in the future?
- How do I learn more about data mining, machine learning, and AI?
- What other topics are relevant to data science beyond predictive modeling?
- What jobs out there value the skills I learned in STAT 4710? How do I get those jobs?

What comes next?

- Where are data mining, machine learning, and AI going in the future?
- How do I learn more about data mining, machine learning, and AI?
- What other topics are relevant to data science beyond predictive modeling?
- What jobs out there value the skills I learned in STAT 4710? How do I get those jobs?
- How do I prepare for the impact of AI on data science?

Where are data mining, ML, and AI heading?

Where are data mining, ML, and AI heading?

- Businesses increasingly data driven; data mining and ML will continue becoming increasingly common in marketing, finance, e-commerce, etc.

Where are data mining, ML, and AI heading?

- Businesses increasingly data driven; data mining and ML will continue becoming increasingly common in marketing, finance, e-commerce, etc.
- Regression and tree-based methods will continue to be used for general-purpose data mining; deep learning for images and natural language.

Where are data mining, ML, and AI heading?

- Businesses increasingly data driven; data mining and ML will continue becoming increasingly common in marketing, finance, e-commerce, etc.

- Regression and tree-based methods will continue to be used for general-purpose data mining; deep learning for images and natural language.

- Datasets are becoming increasingly bigger, so more emphasis will be placed on large-scale/cloud computation and parallelization.

Where are data mining, ML, and AI heading?

- Businesses increasingly data driven; data mining and ML will continue becoming increasingly common in marketing, finance, e-commerce, etc.
- Regression and tree-based methods will continue to be used for general-purpose data mining; deep learning for images and natural language.
- Datasets are becoming increasingly bigger, so more emphasis will be placed on large-scale/cloud computation and parallelization.
- As deep learning matures, more emphasis will be placed on understanding and interpretation, making it more safe, robust, and fair, developing theory.

Where are data mining, ML, and AI heading?

- Businesses increasingly data driven; data mining and ML will continue becoming increasingly common in marketing, finance, e-commerce, etc.
- Regression and tree-based methods will continue to be used for general-purpose data mining; deep learning for images and natural language.
- Datasets are becoming increasingly bigger, so more emphasis will be placed on large-scale/cloud computation and parallelization.
- As deep learning matures, more emphasis will be placed on understanding and interpretation, making it more safe, robust, and fair, developing theory.
- In the short term, AI will automate routine tasks, and serve as an increasingly capable assistant. In the medium-to-longer term, hard to predict!

Learning more about data mining, ML, and AI

	Computation	Theory	Artificial intelligence
Basic			
Advanced			

Learning more about data mining, ML, and AI

	Computation	Theory	Artificial intelligence
Basic	• Python programming (CIS1920, STAT4770)		
Advanced			

Learning more about data mining, ML, and AI

	Computation	Theory	Artificial intelligence
Basic	• Python programming (CIS1920, STAT4770)		
Advanced	• Large-scale computing (NETS2120,CIS5450) • Databases (CIS4500, OID 3150)		

Learning more about data mining, ML, and AI

	Computation	Theory	Artificial intelligence
Basic	• Python programming (CIS1920, STAT4770)	• Probability (STAT4300, ESE3010) • Linear algebra (MATH2400, MATH3120) • Calculus (MATH1140, MATH1150)	
Advanced	• Large-scale computing (NETS2120,CIS5450) • Databases (CIS4500, OID 3150)		

Learning more about data mining, ML, and AI

	Computation	Theory	Artificial intelligence
Basic	• Python programming (CIS1920, STAT4770)	• Probability (STAT4300, ESE3010) • Linear algebra (MATH2400, MATH3120) • Calculus (MATH1140, MATH1150)	
Advanced	• Large-scale computing (NETS2120,CIS5450) • Databases (CIS4500, OID 3150)	• Mathy machine learning (CIS5200,ESE5450) • Optimization (STAT 4810, CIS5150, ESE5040, ESE6050)	

Learning more about data mining, ML, and AI

	Computation	Theory	Artificial intelligence
Basic	• Python programming (CIS1920, STAT4770)	• Probability (STAT4300, ESE3010) • Linear algebra (MATH2400, MATH3120) • Calculus (MATH1140, MATH1150)	
Advanced	• Large-scale computing (NETS2120,CIS5450) • Databases (CIS4500, OID 3150)	• Mathy machine learning (CIS5200,ESE5450) • Optimization (STAT 4810, CIS5150, ESE5040, ESE6050)	• Artificial intelligence (CIS 5210) • Deep learning (CIS5220) • Computer vision (CIS6800) • NLP (CIS5300)

Beyond predictive modeling

Beyond predictive modeling

Not every problem in the world can be solved by machine learning! 🤯

Beyond predictive modeling

Not every problem in the world can be solved by machine learning! 🤯

Statistical inference versus prediction: understanding the world is different from predicting it. E.g., science is driven by understanding rather than prediction.

Beyond predictive modeling

Not every problem in the world can be solved by machine learning! 🤯

Statistical inference versus prediction: understanding the world is different from predicting it. E.g., science is driven by understanding rather than prediction.

Causal inference: Cause and effect is the ultimate question in science and policy. Decision makers (e.g. the FDA) need to estimate the effects of different policies.

Beyond predictive modeling

Not every problem in the world can be solved by machine learning! 🤯

Statistical inference versus prediction: understanding the world is different from predicting it. E.g., science is driven by understanding rather than prediction.

Causal inference: Cause and effect is the ultimate question in science and policy. Decision makers (e.g. the FDA) need to estimate the effects of different policies.

Quantifying uncertainty: If a neural network predicts there is no pedestrian ahead with probability 0.99, what does this mean? Can we make a statistical guarantee?

Beyond predictive modeling

Not every problem in the world can be solved by machine learning!

Statistical inference versus prediction: understanding the world is different from predicting it. E.g., science is driven by understanding rather than prediction.

Causal inference: Cause and effect is the ultimate question in science and policy. Decision makers (e.g. the FDA) need to estimate the effects of different policies.

Quantifying uncertainty: If a neural network predicts there is no pedestrian ahead with probability 0.99, what does this mean? Can we make a statistical guarantee?

Robustness and safety: Will the self-driving car recognize a pedestrian if she is holding an umbrella? In what sense can we ensure the robustness of an algorithm?

Beyond predictive modeling

Not every problem in the world can be solved by machine learning! 🤯

Statistical inference versus prediction: understanding the world is different from predicting it. E.g., science is driven by understanding rather than prediction.

Causal inference: Cause and effect is the ultimate question in science and policy. Decision makers (e.g. the FDA) need to estimate the effects of different policies.

Quantifying uncertainty: If a neural network predicts there is no pedestrian ahead with probability 0.99, what does this mean? Can we make a statistical guarantee?

Robustness and safety: Will the self-driving car recognize a pedestrian if she is holding an umbrella? In what sense can we ensure the robustness of an algorithm?

Fairness: In what sense can a prediction rule be considered fair? How can we assure that our predictive rules live up to this standard?

Data science jobs

Data Scientist: The Sexiest Job of the 21st Century

Harvard Business Review

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and D.J. Patil

From the Magazine (October 2012)

glassdoor

50 Best Jobs in America for 2022

	Job Title	Median Base Salary	Job Satisfaction	Job Openings
#1	Enterprise Architect	$144,997	4.1/5	14,021
#2	Full Stack Engineer	$101,794	4.3/5	11,252
#3	Data Scientist	$120,000	4.1/5	10,071
#4	Devops Engineer	$120,095	4.2/5	8,548
#5	Strategy Manager	$140,000	4.2/5	6,977

What does a Data Scientist do?

Data scientists utilize their analytical, statistical, and programming skills to collect, analyze, and interpret large data sets. They then use this information to develop data-driven solutions to difficult business challenges. Data scientists commonly have a bachelor's degree in statistics, math, computer science, or economics. Data scientists have a wide range of technical...

Read More

Average Years of Experience

- 0 - 1: 8%
- 2 - 4: 80%
- 5 - 7: 6%
- 8+: 6%

Common Skill Sets

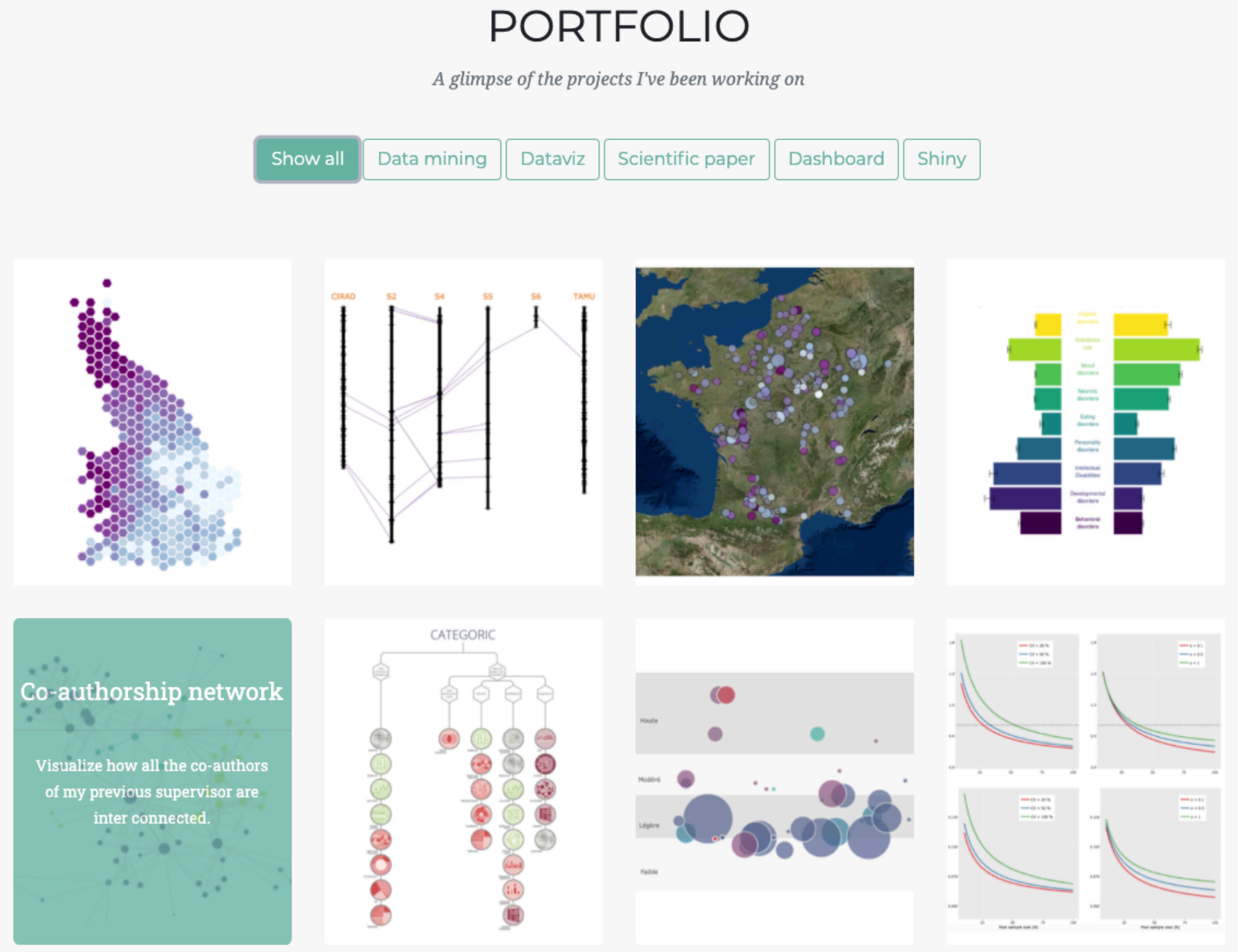
- Machine Learning
- Python
- Hadoop SPARK
- SQL
- Statistics
- Natural Language Processing
- Algorithms
- Programming Languages

How to get a data science job?

- Learn the skills through classes or on your own.
- Build your skills through data science projects.
- Share your work by posting code on Github and making a portfolio of your projects.
- Apply to internships to gain data science experience.

How to get a data science job?

- Learn the skills through classes or on your own.
- Build your skills through data science projects.
- Share your work by posting code on Github and making a portfolio of your projects.
- Apply to internships to gain data science experience.



https://www.yan-holtz.com/

How to prepare for the impact of AI?

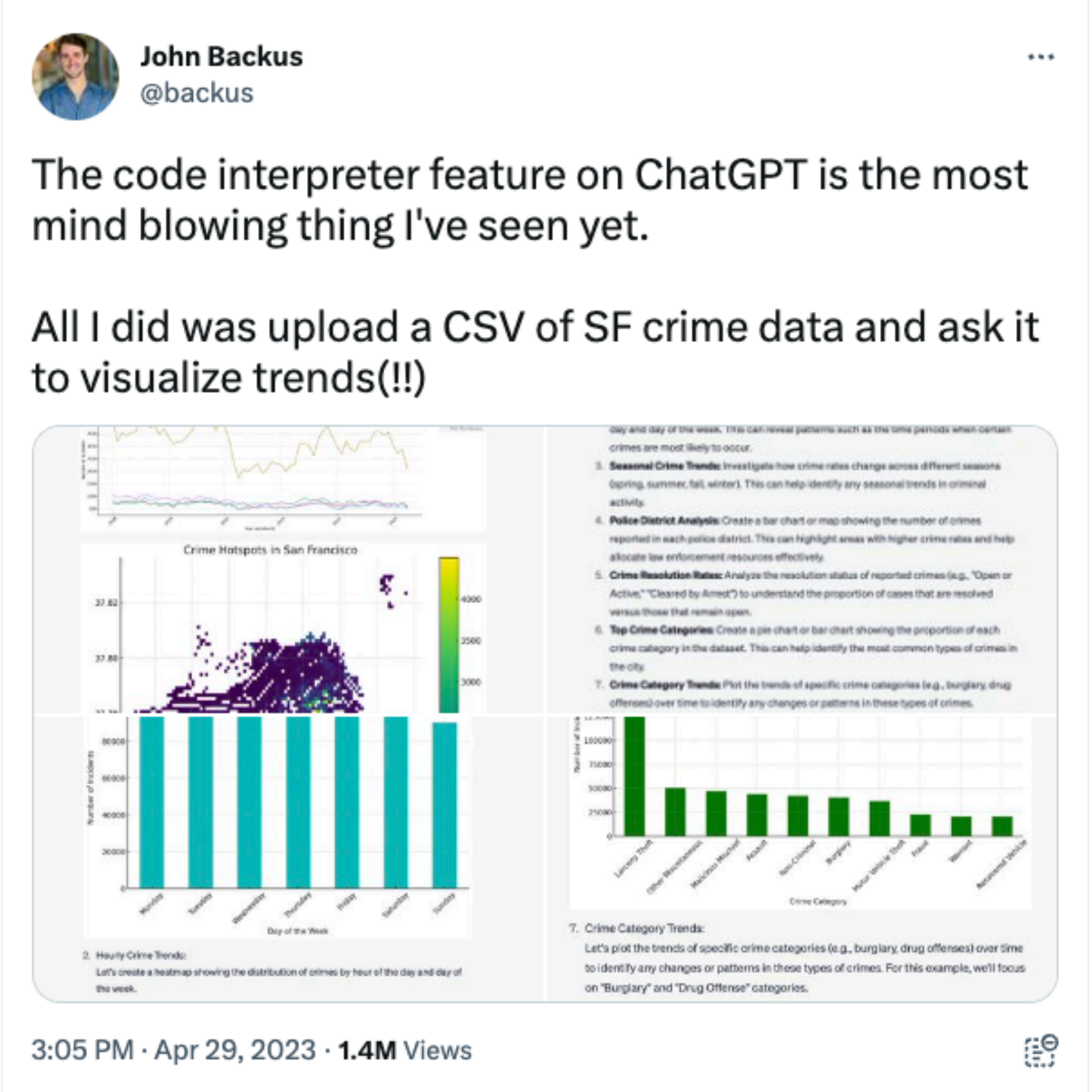
- Learn about the latest AI technologies and capabilities, e.g. code interpreter and retrieval-augmented generation.

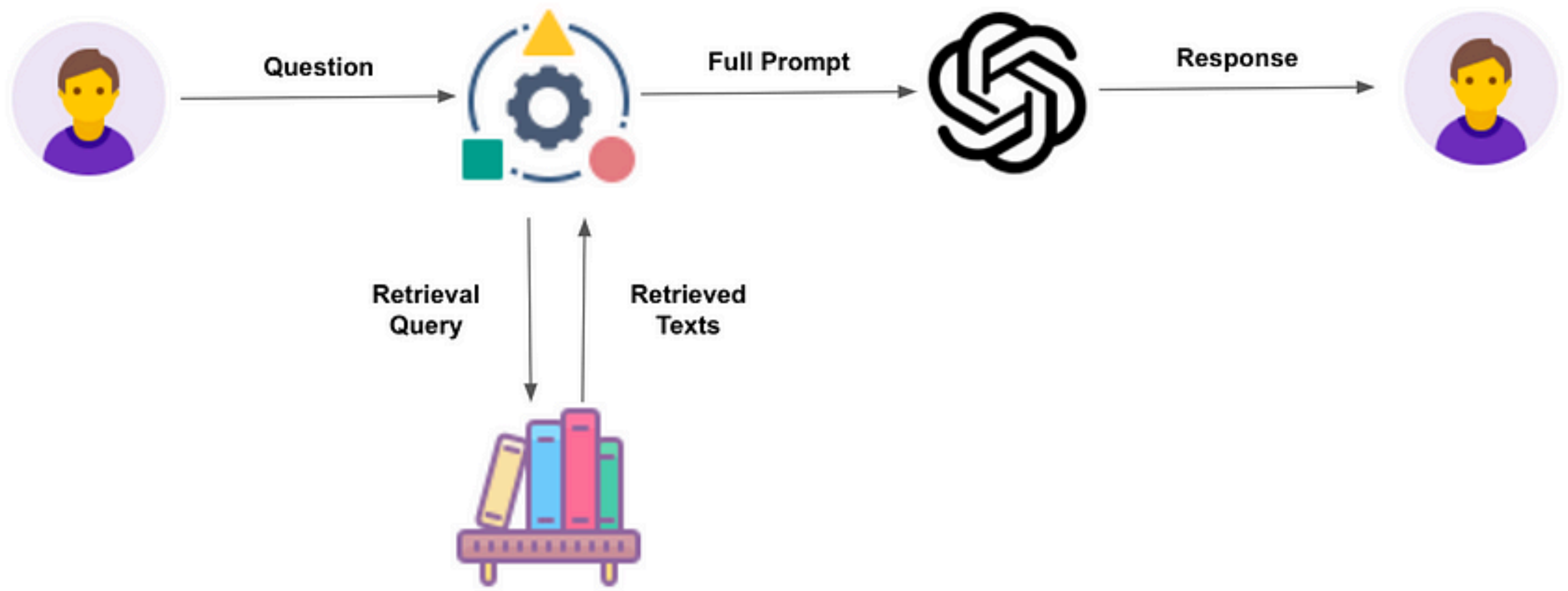
How to prepare for the impact of AI?

- Learn about the latest AI technologies and capabilities, e.g. code interpreter and retrieval-augmented generation.

How to prepare for the impact of AI?

- Learn about the latest AI technologies and capabilities, e.g. code interpreter and retrieval-augmented generation.



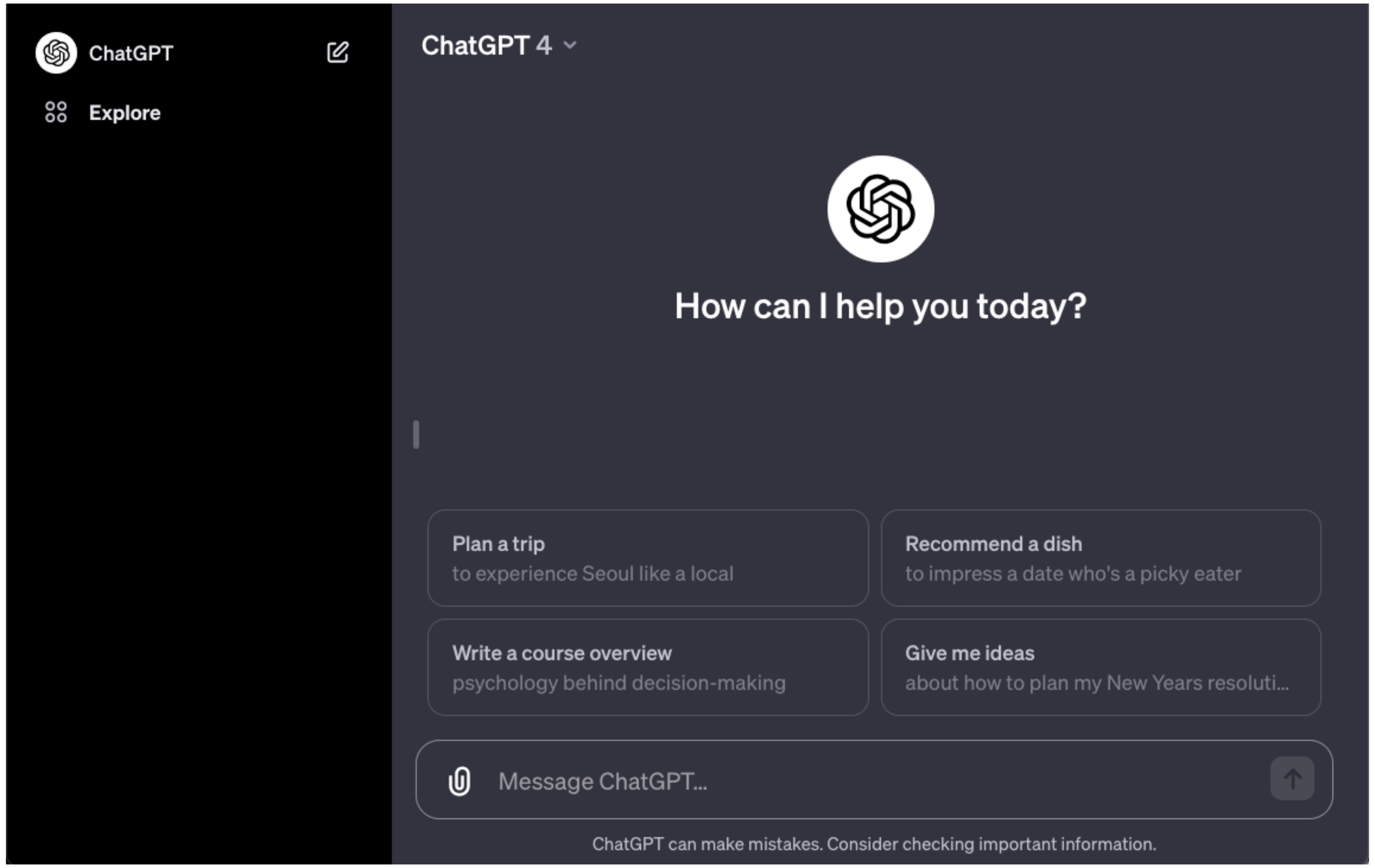


How to prepare for the impact of AI?

- Play around with GUI-based AI tools such as ChatGPT to better understand their strengths and weaknesses, and how prompts affect quality of outputs.

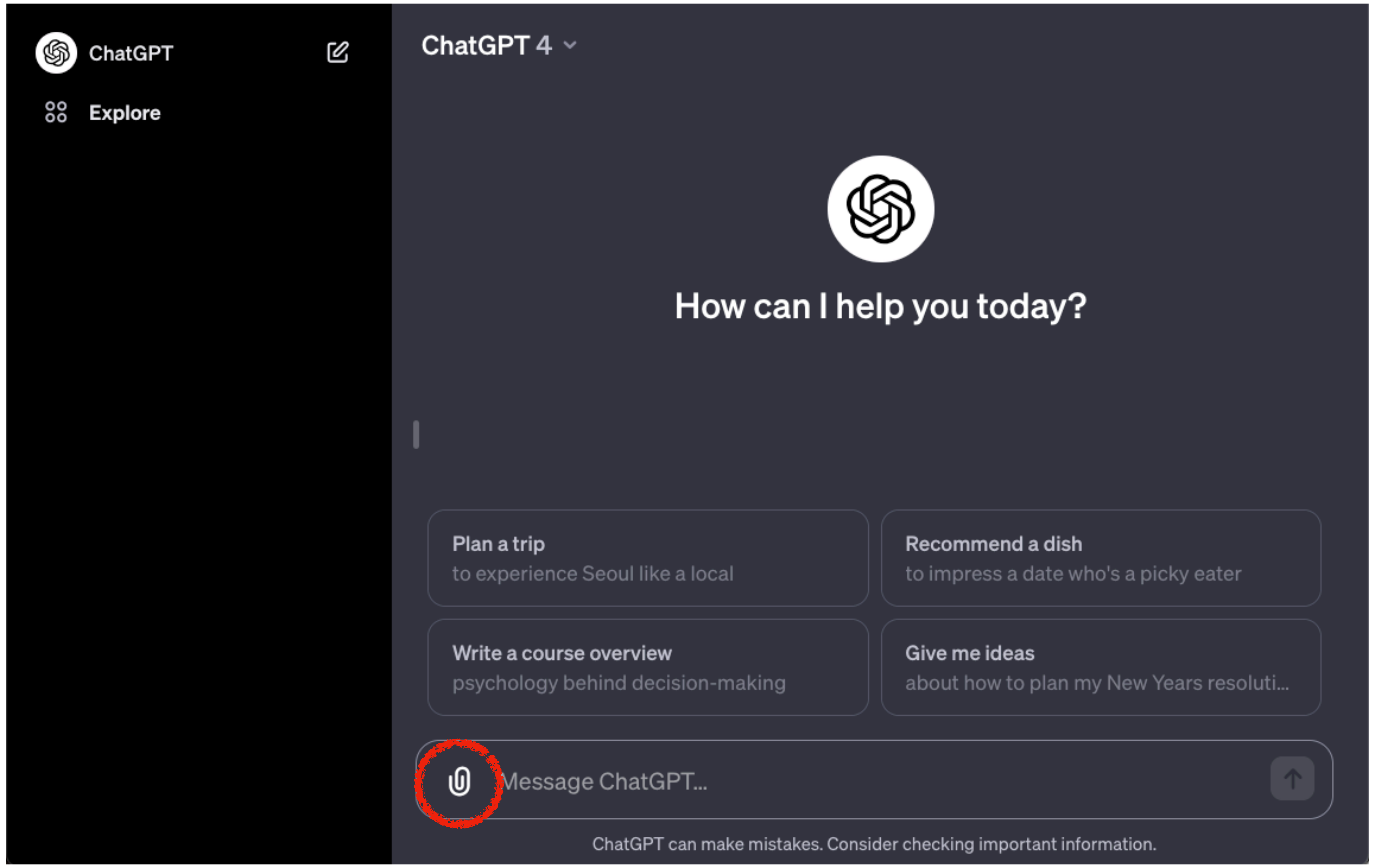
How to prepare for the impact of AI?

- Play around with GUI-based AI tools such as ChatGPT to better understand their strengths and weaknesses, and how prompts affect quality of outputs.



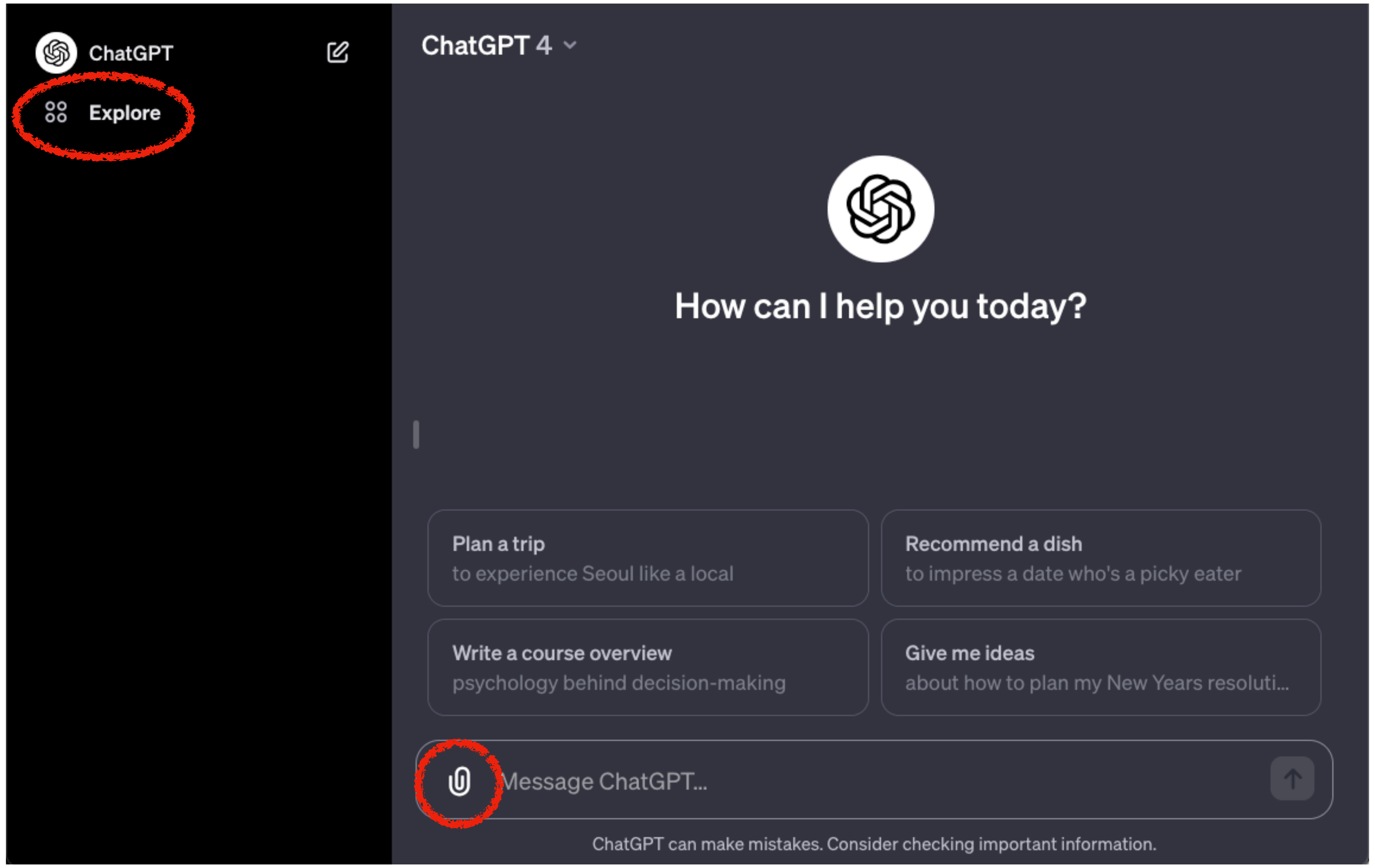
How to prepare for the impact of AI?

- Play around with GUI-based AI tools such as ChatGPT to better understand their strengths and weaknesses, and how prompts affect quality of outputs.



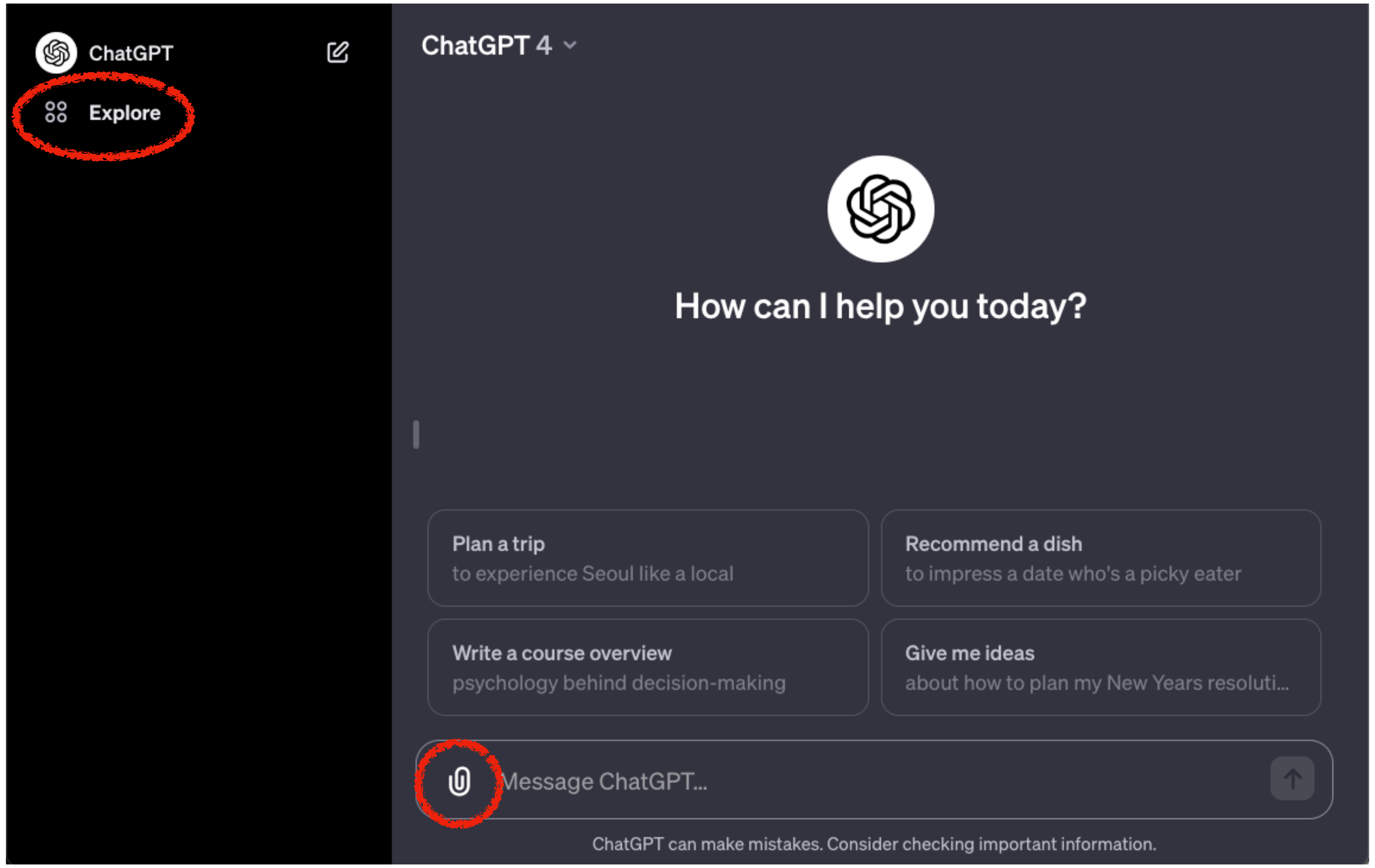
How to prepare for the impact of AI?

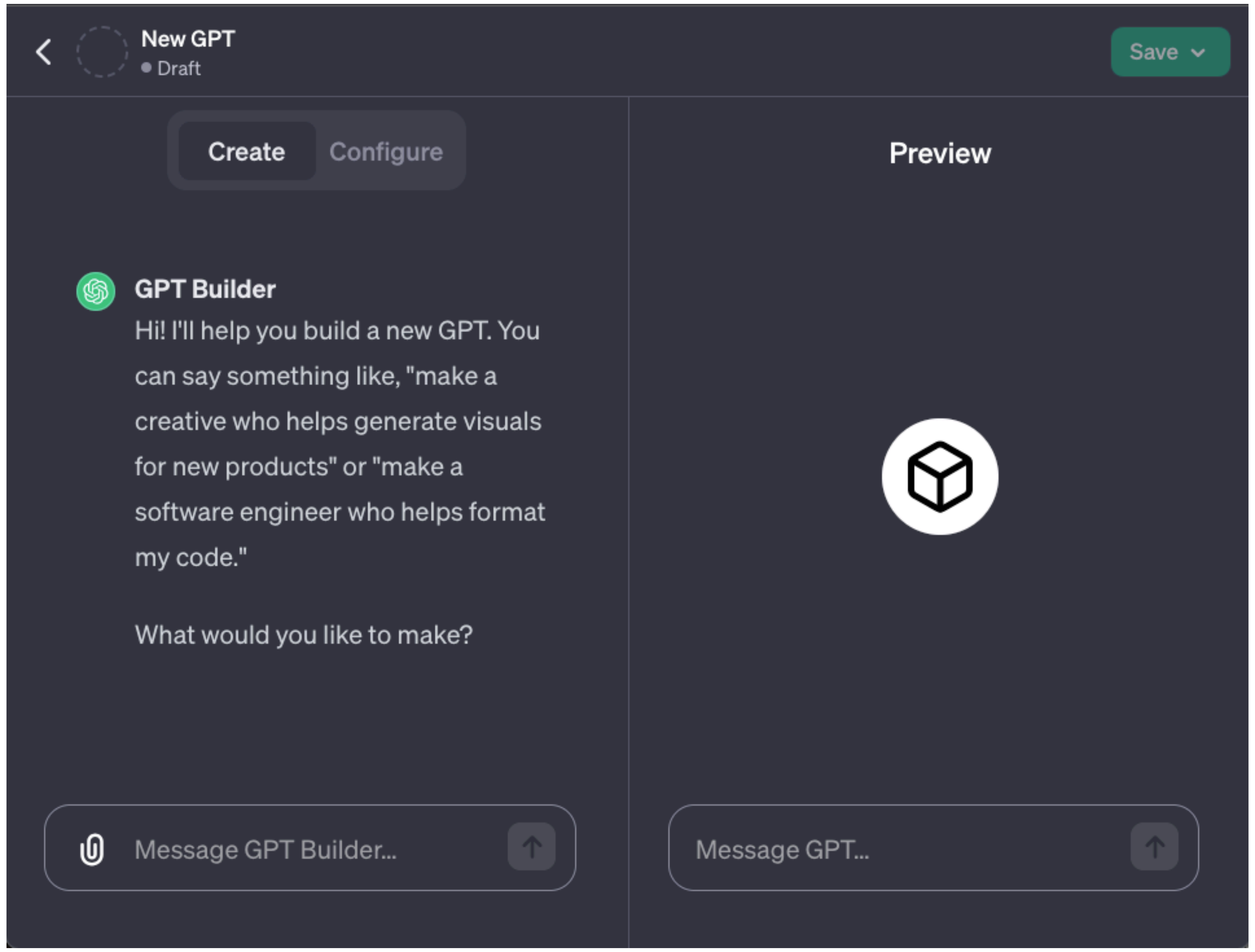
- Play around with GUI-based AI tools such as ChatGPT to better understand their strengths and weaknesses, and how prompts affect quality of outputs.



How to prepare for the impact of AI?

- Play around with GUI-based AI tools such as ChatGPT to better understand their strengths and weaknesses, and how prompts affect quality of outputs.





How to prepare for the impact of AI?

How to prepare for the impact of AI?

- Build on top of existing AI tools:

How to prepare for the impact of AI?

- Build on top of existing AI tools:
  - Build a chatbot based on retrieval-augmented generation.

How to prepare for the impact of AI?

- Build on top of existing AI tools:
  - Build a chatbot based on retrieval-augmented generation.
    (Build an AI TA by equipping LLM with course notes.)

How to prepare for the impact of AI?

- Build on top of existing AI tools:
  - Build a chatbot based on retrieval-augmented generation.
    (Build an AI TA by equipping LLM with course notes.)
  - Fine-tune an existing large language model.

How to prepare for the impact of AI?

- Build on top of existing AI tools:
  - Build a chatbot based on retrieval-augmented generation.
    (Build an AI TA by equipping LLM with course notes.)
  - Fine-tune an existing large language model.
    (Build sentiment classifier by fine-tuning an LLM on Rotten Tomatoes data.)

How to prepare for the impact of AI?

- Build on top of existing AI tools:
  - Build a chatbot based on retrieval-augmented generation.
    (Build an AI TA by equipping LLM with course notes.)
  - Fine-tune an existing large language model.
    (Build sentiment classifier by fine-tuning an LLM on Rotten Tomatoes data.)
  - Build an application on top of an existing large language model.

How to prepare for the impact of AI?

- Build on top of existing AI tools:
  - Build a chatbot based on retrieval-augmented generation.
    (Build an AI TA by equipping LLM with course notes.)
  - Fine-tune an existing large language model.
    (Build sentiment classifier by fine-tuning an LLM on Rotten Tomatoes data.)
  - Build an application on top of an existing large language model.
    (E.g. Build a simple app for language learning that simulates conversations in different languages, provides grammar corrections, and offers cultural insights.)

How to prepare for the impact of AI?

- Build on top of existing AI tools:
  - Build a chatbot based on retrieval-augmented generation.
    (Build an AI TA by equipping LLM with course notes.)
  - Fine-tune an existing large language model.
    (Build sentiment classifier by fine-tuning an LLM on Rotten Tomatoes data.)
  - Build an application on top of an existing large language model.
    (E.g. Build a simple app for language learning that simulates conversations in different languages, provides grammar corrections, and offers cultural insights.)

How to prepare for the impact of AI?

- Build on top of existing AI tools:
  - Build a chatbot based on retrieval-augmented generation.
    (Build an AI TA by equipping LLM with course notes.)
  - Fine-tune an existing large language model.
    (Build sentiment classifier by fine-tuning an LLM on Rotten Tomatoes data.)
  - Build an application on top of an existing large language model.
    (E.g. Build a simple app for language learning that simulates conversations in different languages, provides grammar corrections, and offers cultural insights.)

See OpenAI developer platform, Hugging Face, and LangChain.