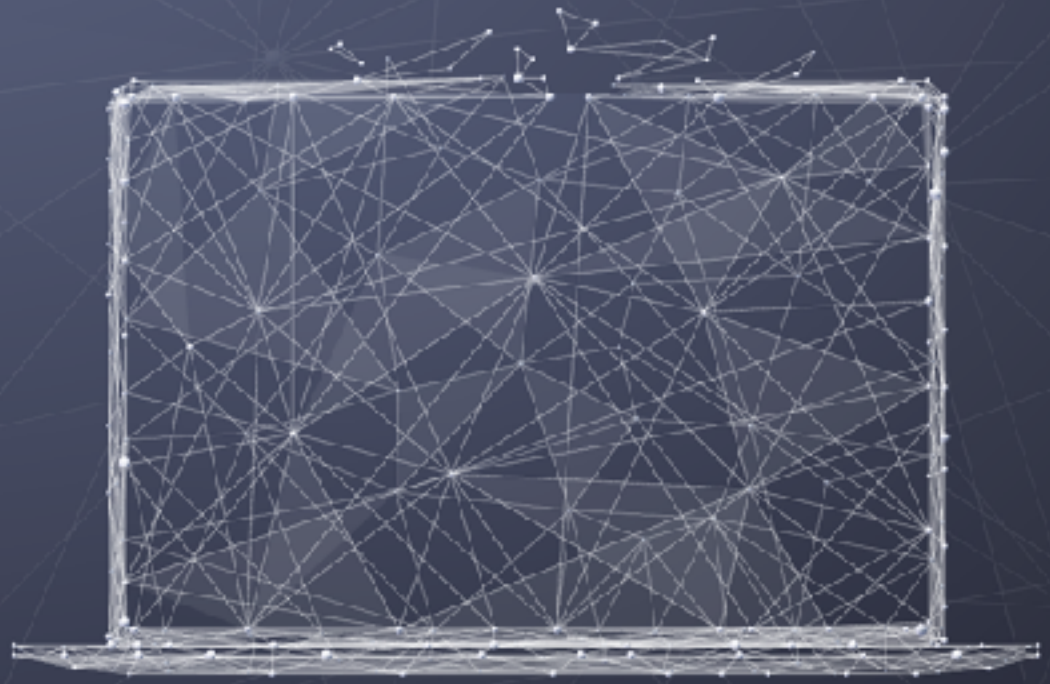
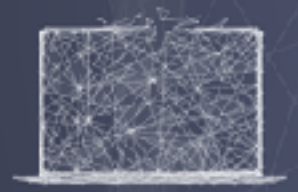
Data Science
Foundations of Decision Making

Predictive modeling



PURDUE UNIVERSITY | College of Science



The Data Analysis Pipeline

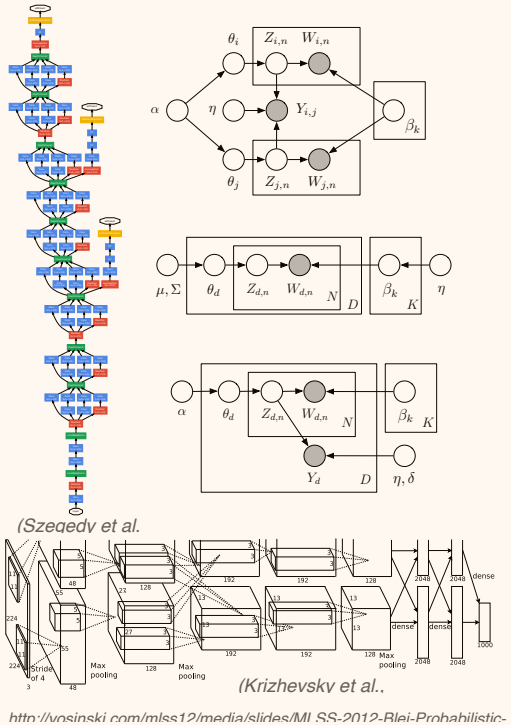
PURDUE UNIVERSITY | College of Science

Source: Jiasen Yang

The Data Analysis Pipeline

https://cs.stanford.edu/people/karpathy/cnnembed/
https://commons.wikimedia.org/wiki/File:Protein_GC_PDB_1j78.png
https://www.b2bmarketing.net/en-gb/resources/blog/network-effect-what-b2b-comms-can-learn-

Collect Data

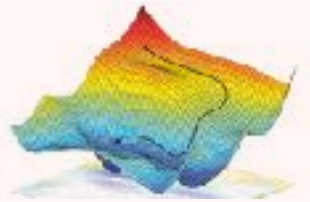


(Szegedy et al.

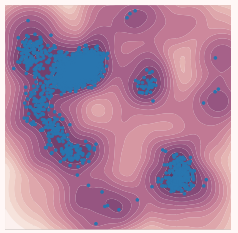
(Krizhevsky et al.,

http://vosinski.com/mlss12/media/slides/MLSS-2012-Blei-Probabilistic-

Design Model



(Li et al.



http://www.sciencemag.org/news/2018/05/

Infer Parameters

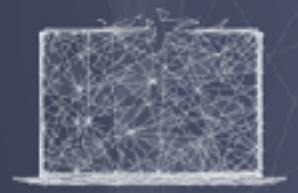
(Adamic & Glance

(Nickel & Kiela

Apply Model

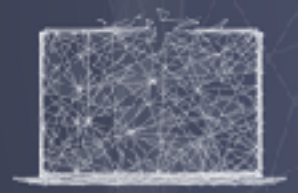
PURDUE UNIVERSITY | College of Science

Source: Jiasen Yang



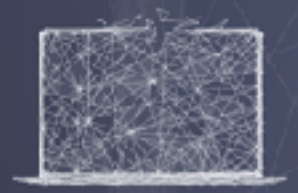
Machine learning tasks

- Supervised learning:
  Given examples with inputs/outputs (X, y), learn a function to map X to y, i.e. f(X)=y
  Goal is to learn f from training data, and evaluate it on new (test) data.
  - Classification: y is discrete (y are class labels, X refers to attributes)
  - Regression: y is continuous
- Unsupervised learning:
  Given examples with only inputs X, learn a function f(X) to simplify the data and map to unknown y
  - Clustering: y is discrete (y are cluster memberships)
  - Matrix factorization: y are continuous (y are embeddings)



Machine learning methods

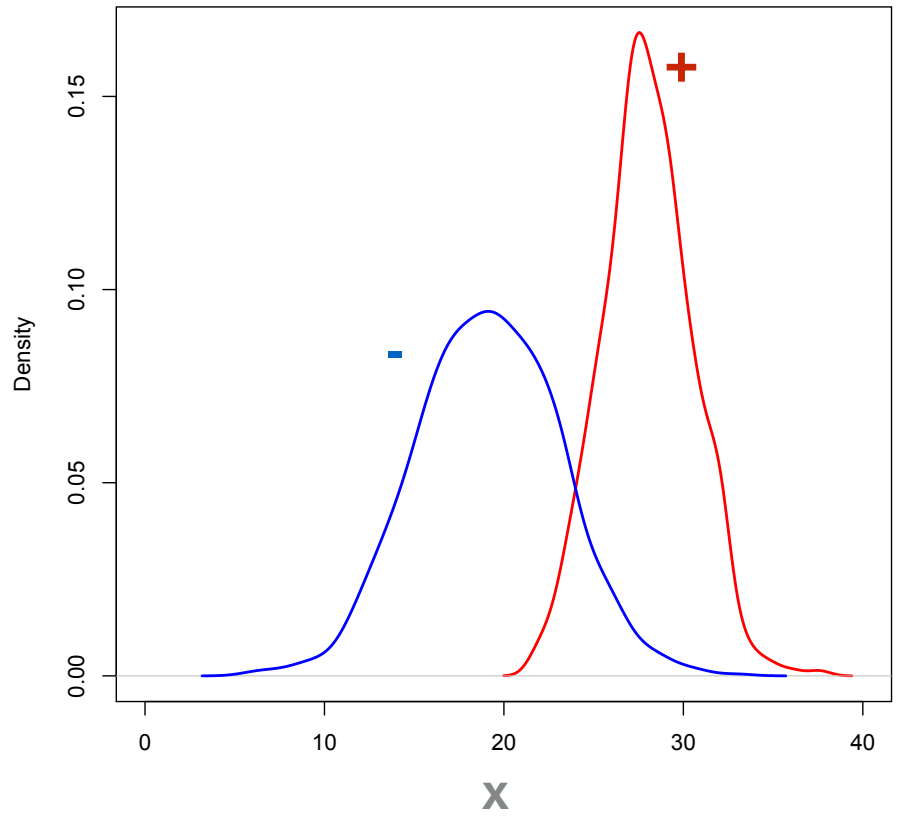
- Method to construct model or patterns from data
- Model space
  - Choice of model representation defines a set of possible models or patterns
- Objective function
  - Associates a numerical value (score) with each member of the set of models/ patterns
- Search technique (i.e., optimization)
  - Defines a method for generating members of the set of models/patterns, determining their score, and identifying the ones with the “best” score



Objective function

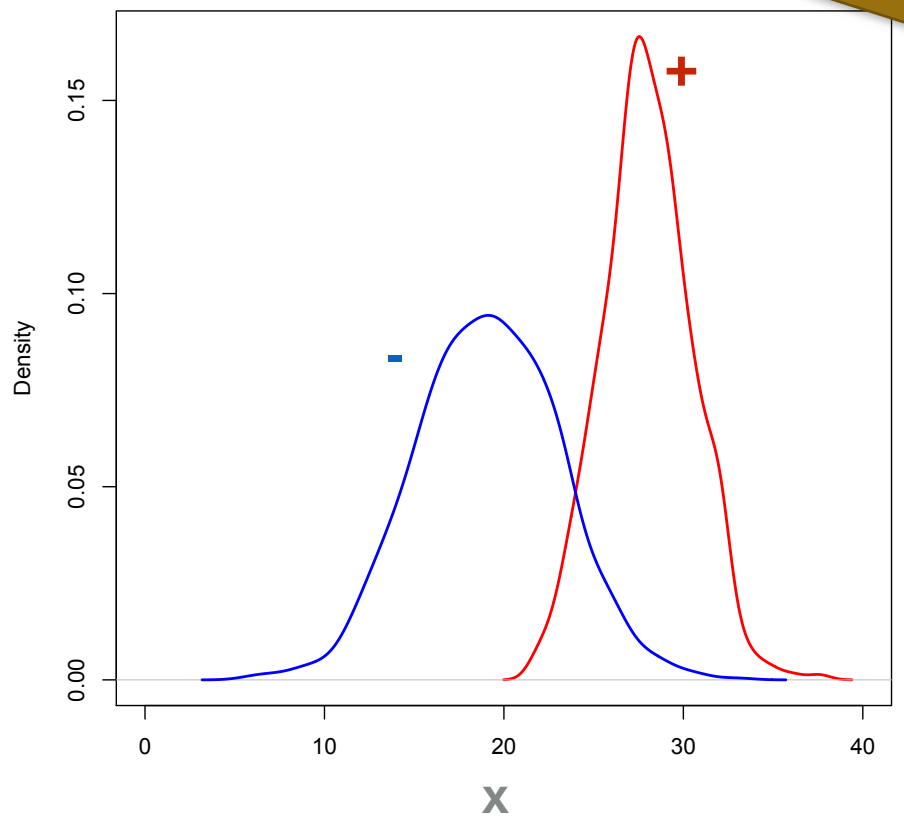
- A numeric score assigned to each possible model in a search space, given a reference/input dataset
  - Used to judge the quality of a particular model for the domain
- Score function are statistics—estimates of a population parameter based on a sample of data
  - Examples: Misclassification, Squared error, Likelihood

Example learning problem



Example learning problem

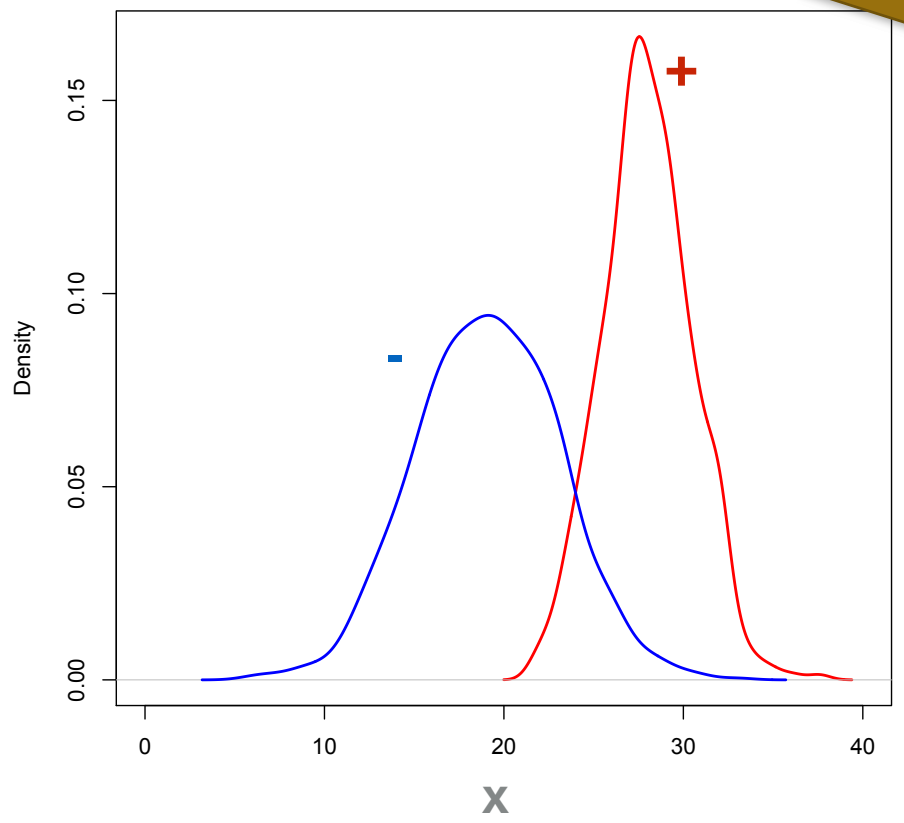
Task: Devise a rule to classify items based on the attribute *X*



PURDUE UNIVERSITY | College of Science

Example learning problem

Model representation:
If-then rules



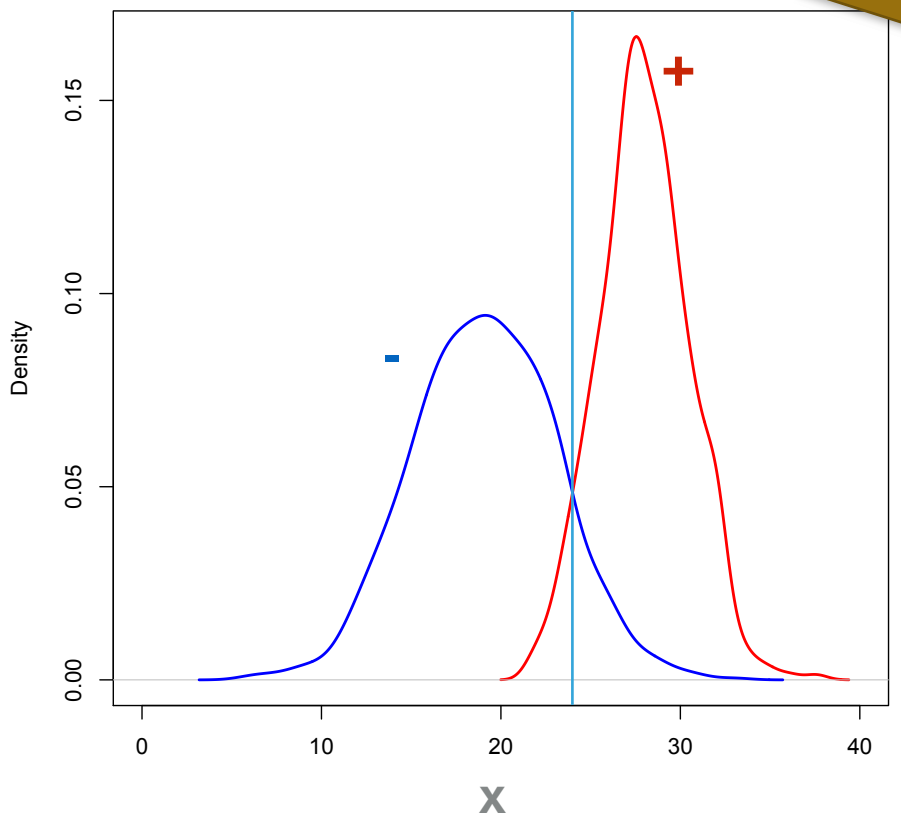
Task: Devise a rule to classify items based on the attribute *X*

PURDUE UNIVERSITY | College of Science

Example learning problem

Model representation:
If-then rules

Example rule:
If x > 25 then + ; Else -



Task: Devise a rule to classify items based on the attribute *X*

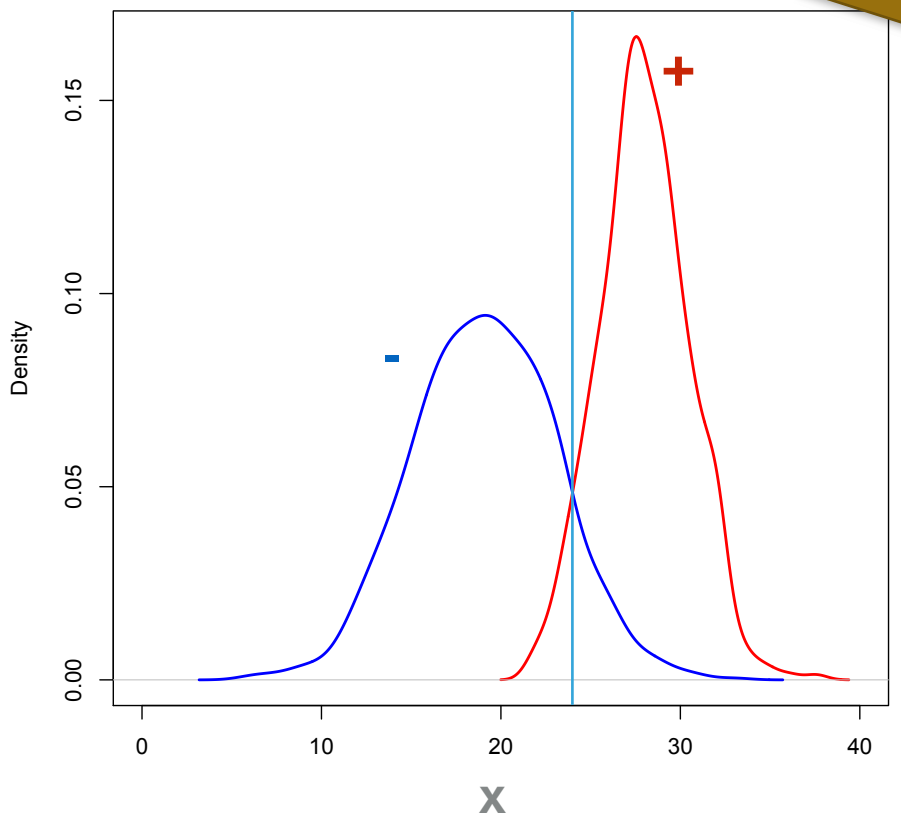
PURDUE UNIVERSITY | College of Science

Example learning problem

Model representation:
If-then rules

Example rule:
If x > 25 then + ; Else -

What is the model space?



Task: Devise a rule to classify items based on the attribute *X*

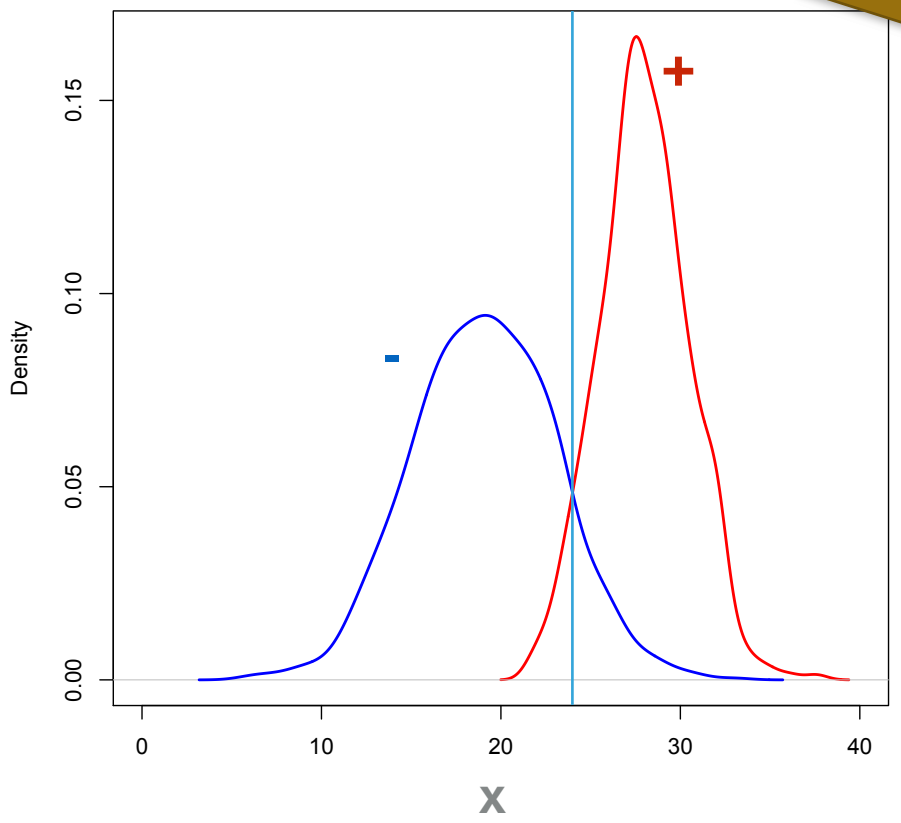
PURDUE UNIVERSITY | College of Science

Example learning problem

Model representation:
If-then rules

Example rule:
If x > 25 then + ; Else -

What is the model space?
All possible thresholds



Task: Devise a rule to classify items based on the attribute *X*

Example learning problem

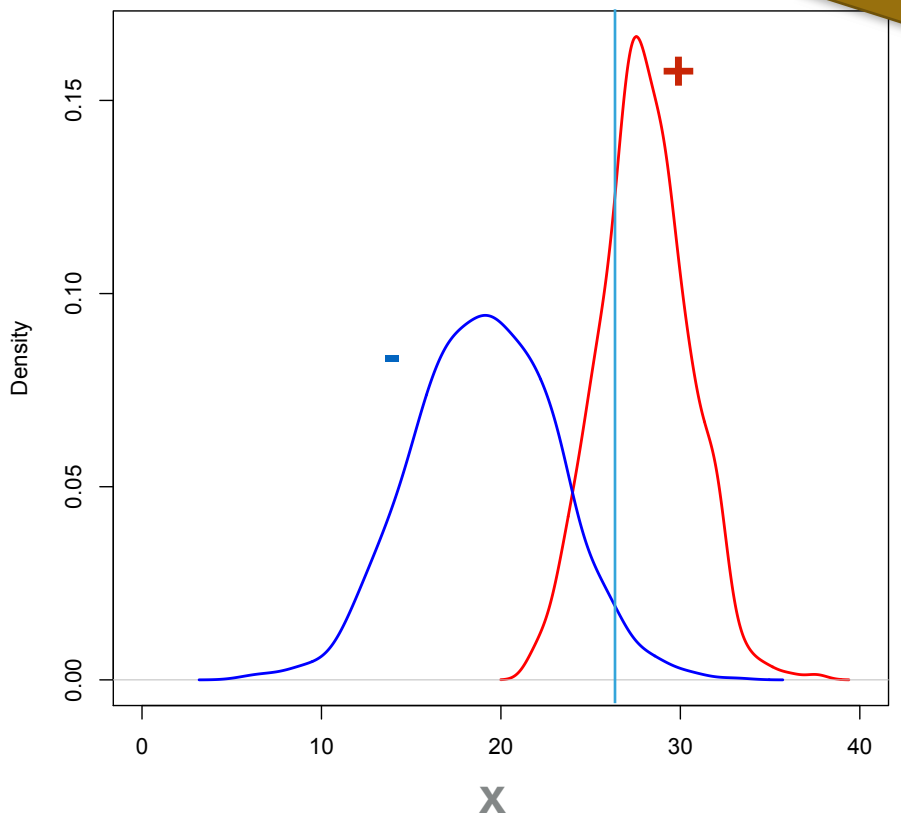
Task: Devise a rule to classify items based on the attribute *X*

Model representation:
If-then rules

Example rule:
If x > 25 then + ; Else -

What is the model space?

All possible thresholds

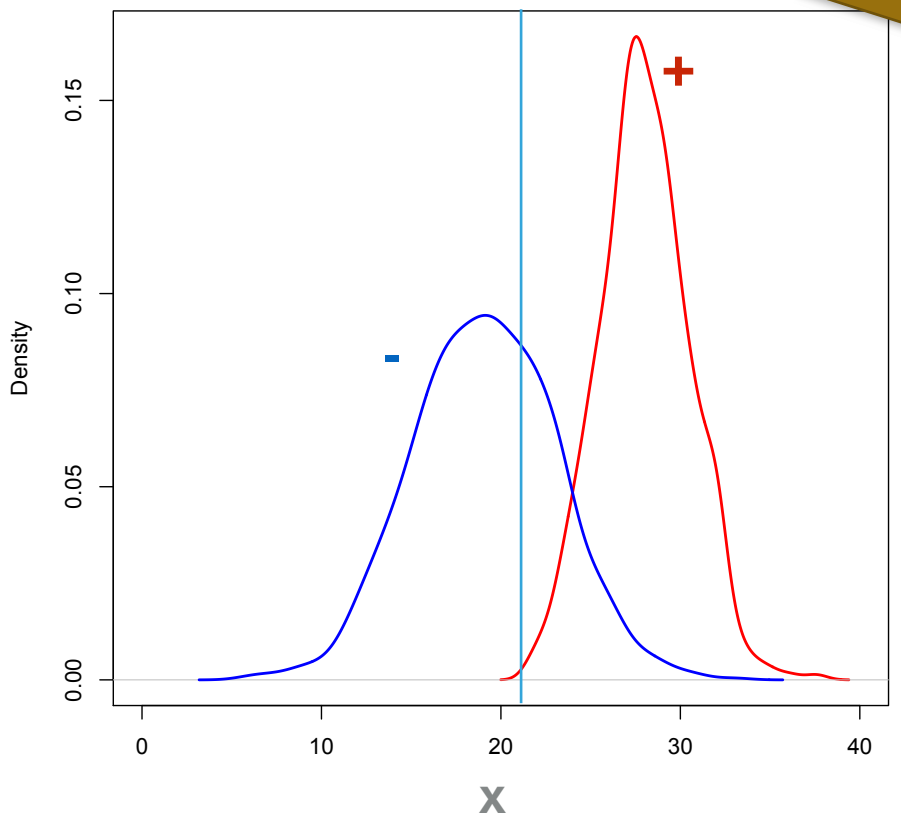


Example learning problem

Model representation:
If-then rules

Example rule:
If x > 25 then + ; Else -

What is the model space?
All possible thresholds



Task: Devise a rule to classify items based on the attribute *X*

Example learning problem

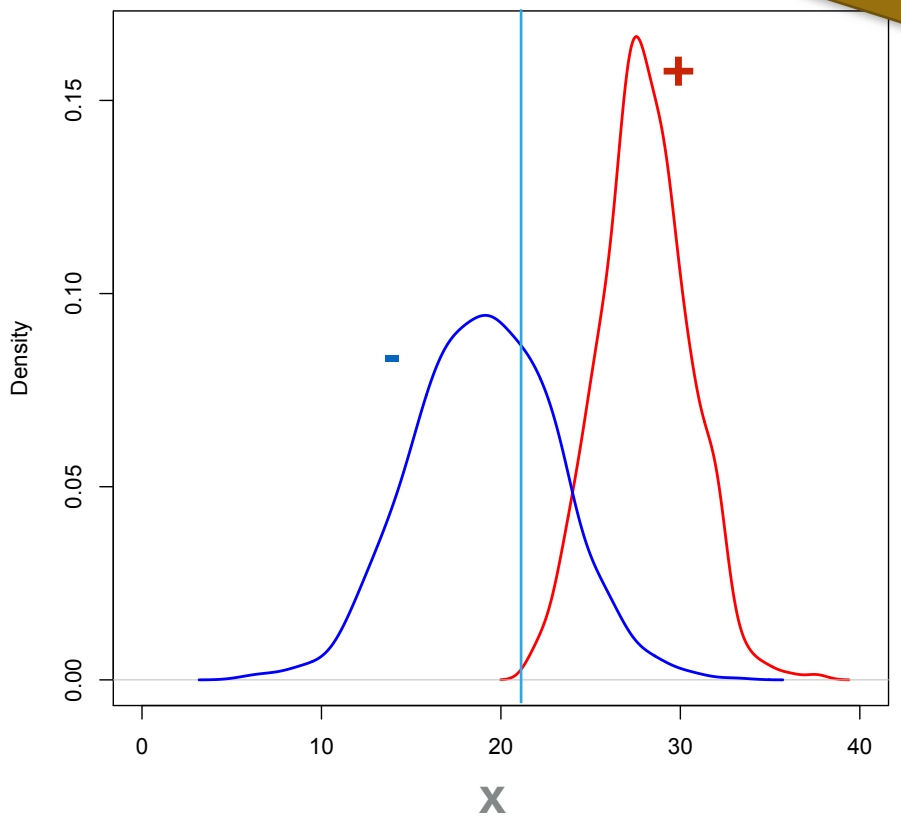
Task: Devise a rule to classify items based on the attribute *X*

Model representation:
If-then rules

Example rule:
If x > 25 then + ; Else -

What is the model space?

All possible thresholds



What is the objective function?

PURDUE UNIVERSITY | College of Science

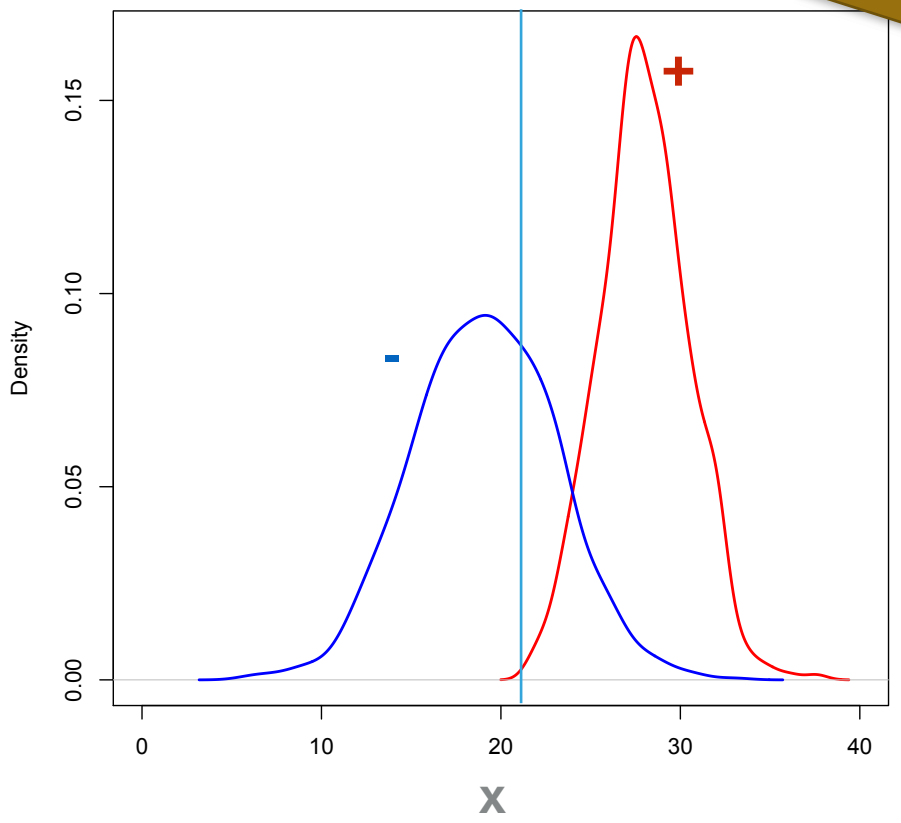
Example learning problem

Task: Devise a rule to classify items based on the attribute *X*

Model representation:
If-then rules

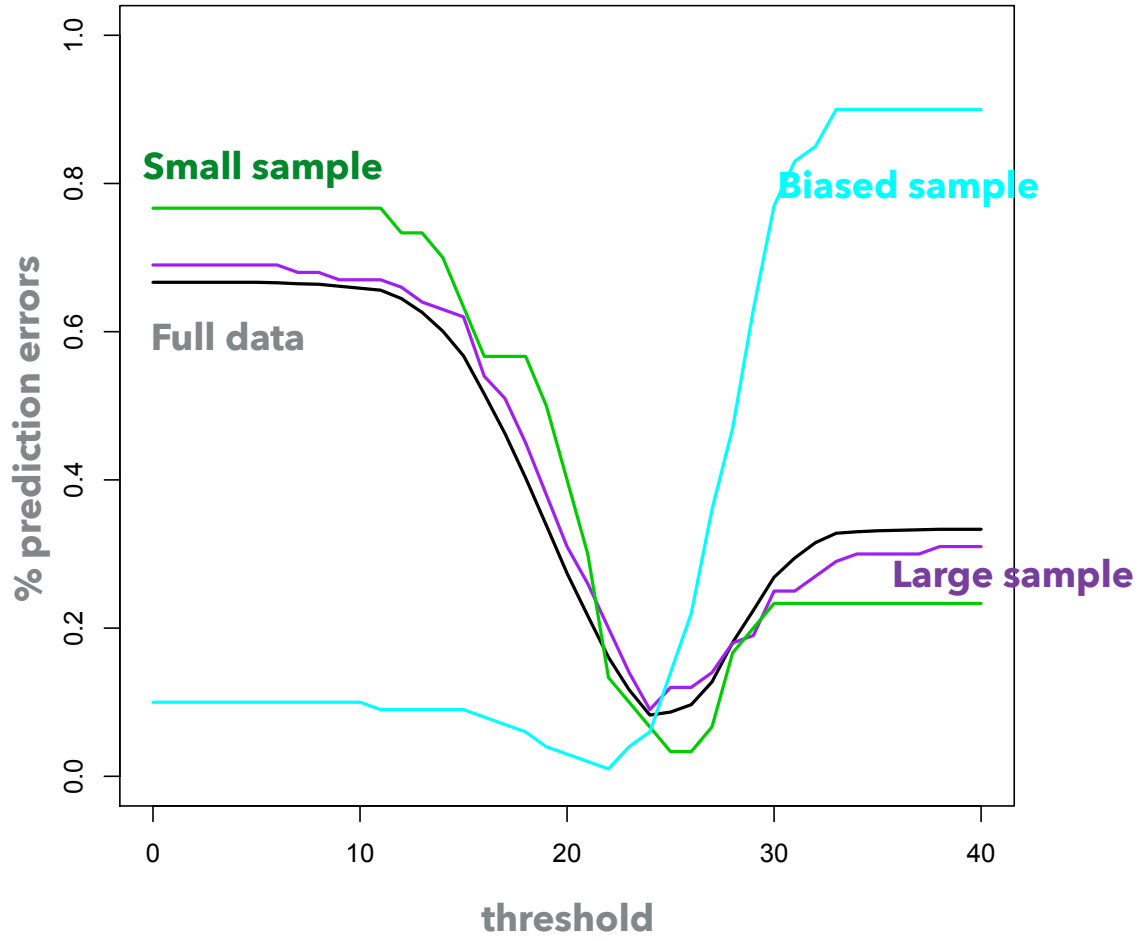
Example rule:
If x > 25 then + ; Else -

What is the model space?
All possible thresholds



What is the objective function?
Prediction error rate

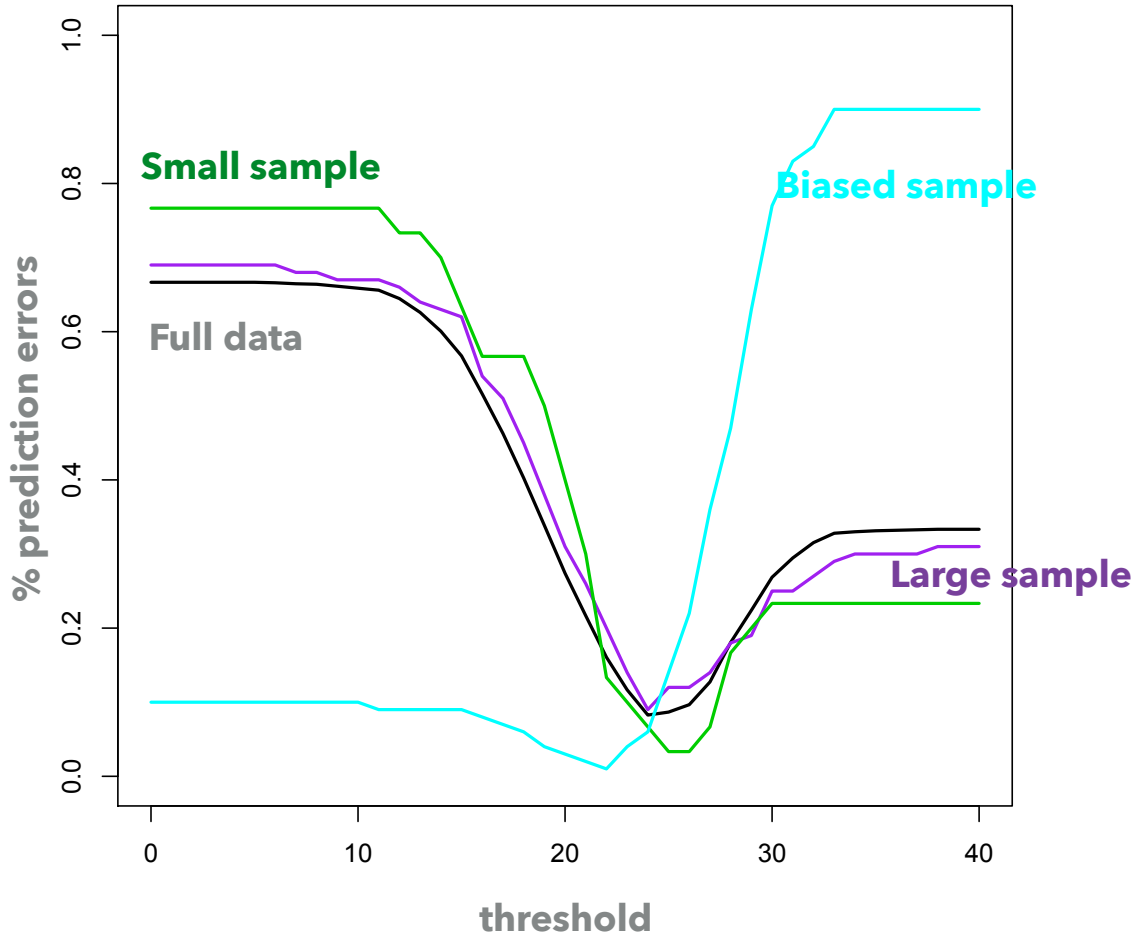
Objective function over model space



PURDUE UNIVERSITY | College of Science

Objective function over model space

Search procedure?

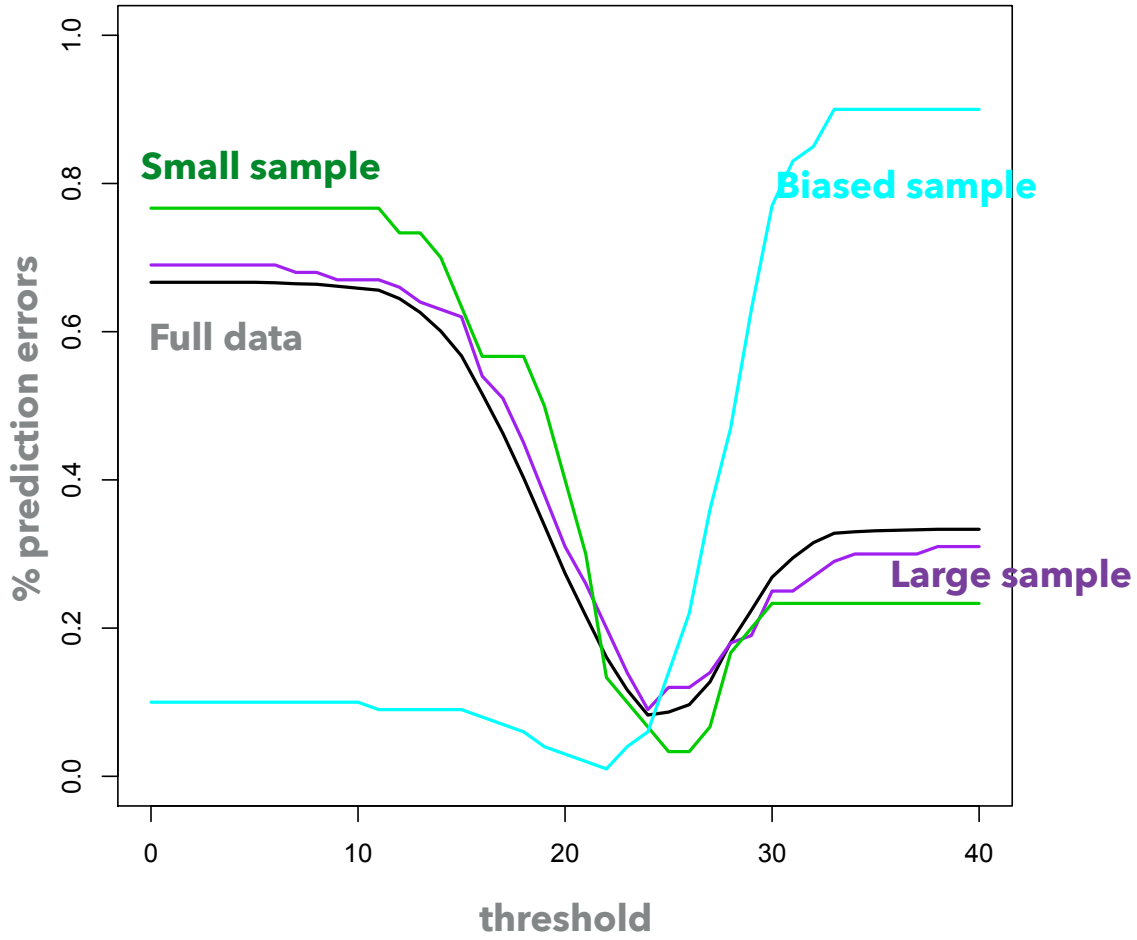


PURDUE UNIVERSITY | College of Science

Objective function over model space

Search procedure?

Try all thresholds, select one with lowest score

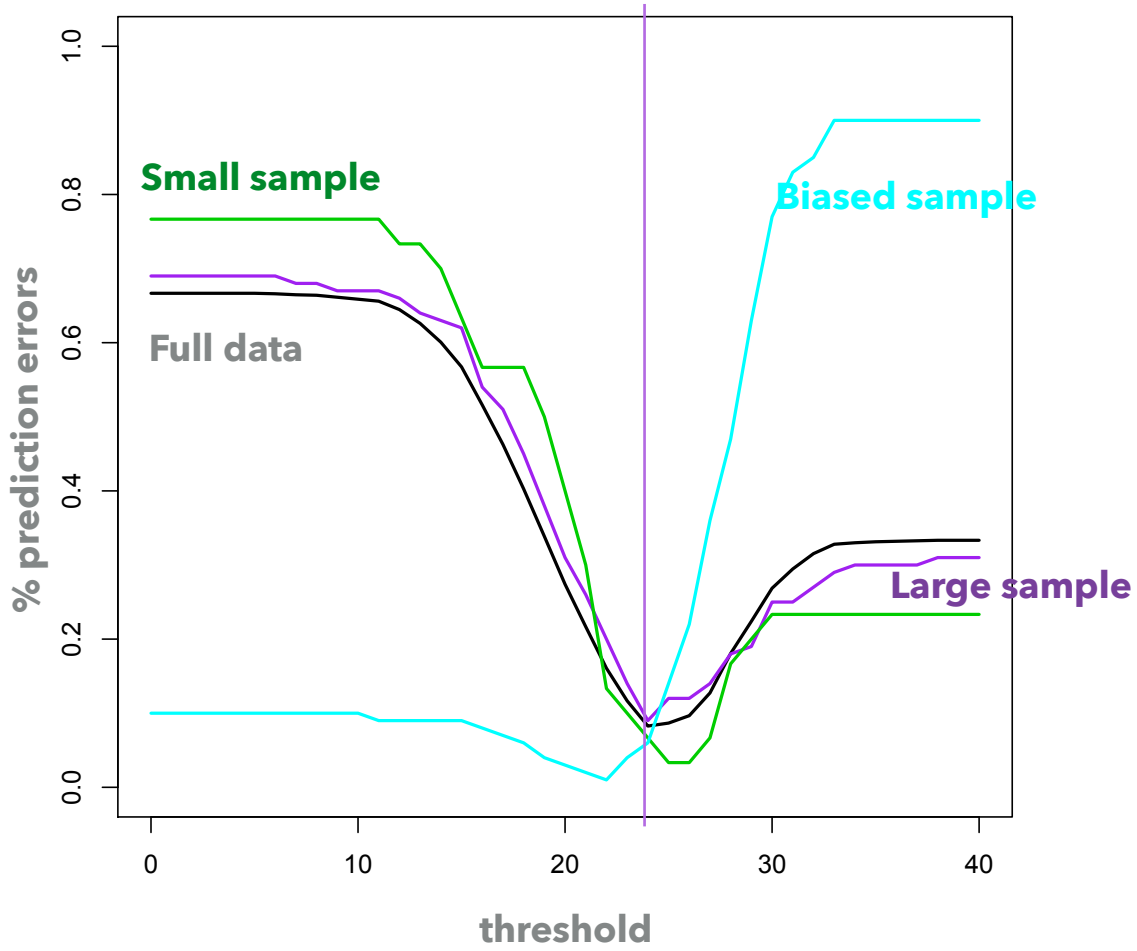


PURDUE UNIVERSITY | College of Science

Objective function over model space

Search procedure?

Try all thresholds, select one with lowest score

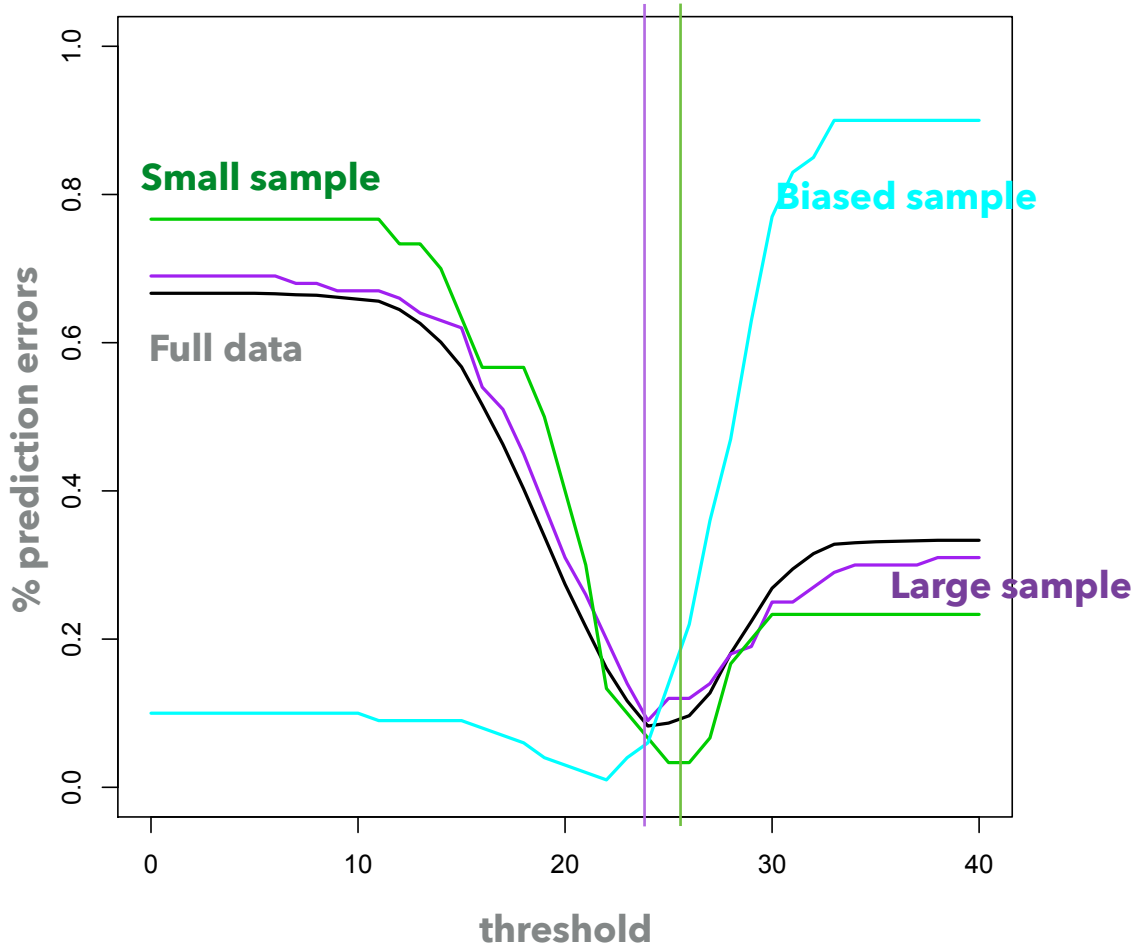


PURDUE UNIVERSITY | College of Science

Objective function over model space

Search procedure?

Try all thresholds, select one with lowest score

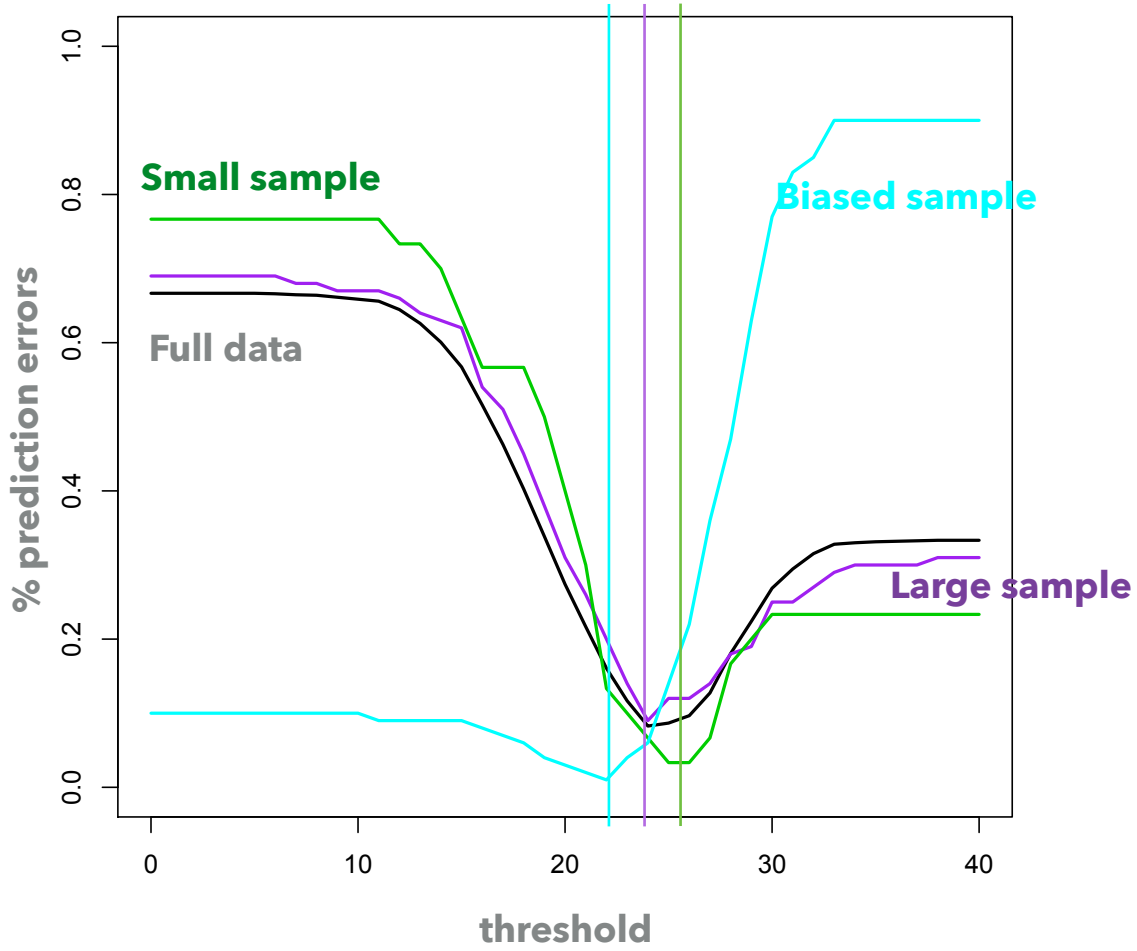


PURDUE UNIVERSITY | College of Science

Objective function over model space

Search procedure?

Try all thresholds, select one with lowest score

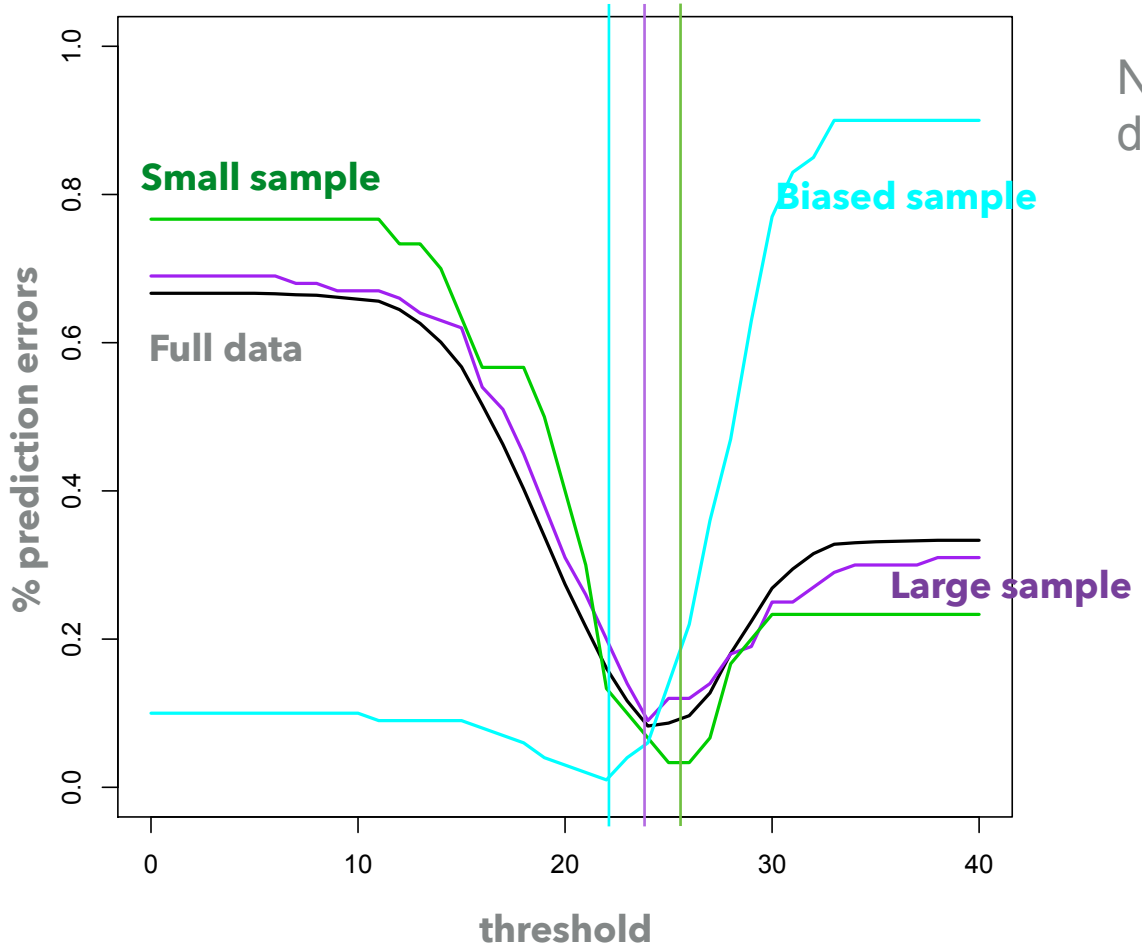


PURDUE UNIVERSITY | College of Science

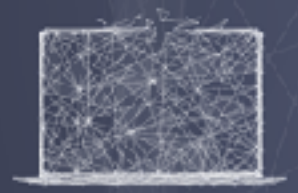
Objective function over model space

Search procedure?

Try all thresholds, select one with lowest score



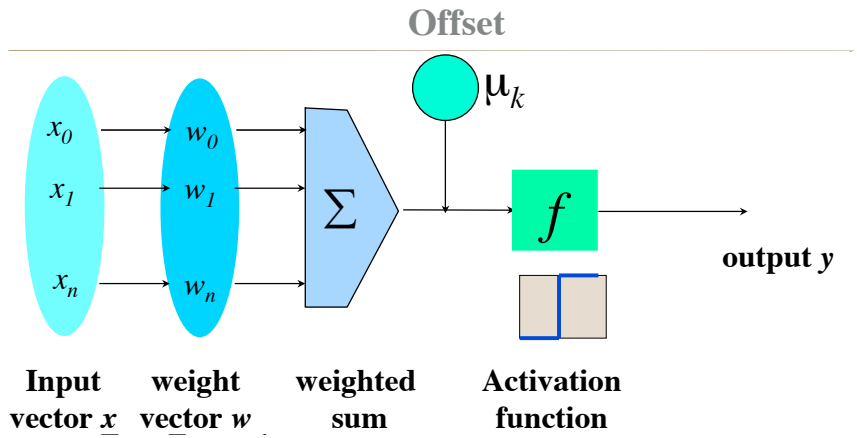
Note: learning result depends on data



Example: Perceptron

Neuron

- First learning algorithm in 1959 (Rosenblatt)
- Perceptron learning rule
- Provide target outputs with inputs for a single neuron
- Incrementally update weights to learn to produce outputs



Perceptron

Model: $f(x) = \sum_{i=1}^{m} w_i x_i + b$

$$y = sign[f(x)]$$

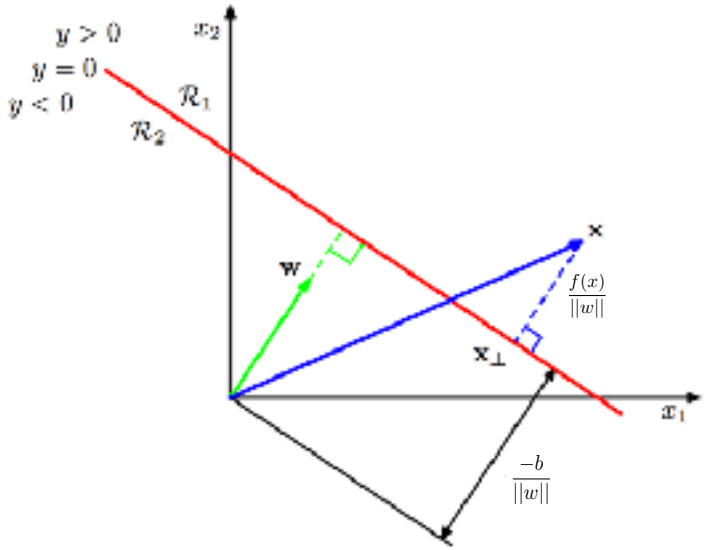


Figure: C. Bishop

Dot product is product of:
(i) projection of x onto w (i.e., ||x|| cos θ), and (ii) the length of w
Dot product is 0 if x is perpendicular to w
Add b, if >0 then positive class, if <0 then negative class

Perceptron

Model: $f(x) = \sum_{i=1}^{m} w_i x_i + b$ *Offset*

$$y = sign[f(x)]$$

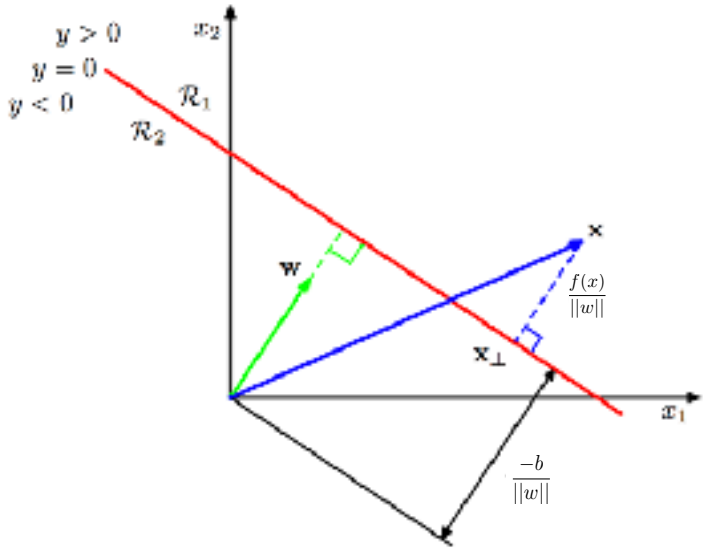


Figure: C. Bishop

Dot product is product of:
(i) projection of x onto w (i.e., ||x|| cos θ), and (ii) the length of w
Dot product is 0 if x is perpendicular to w
Add b, if >0 then positive class, if <0 then negative class

Perceptron

Model: $f(x) = \sum_{i=1}^{m} w_i x_i + b$ *Offset*

$y = sign[f(x)]$

Activation function

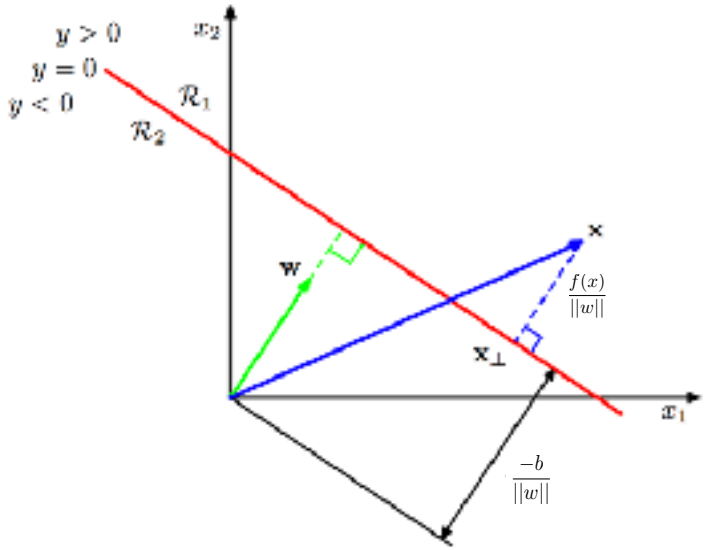


Figure: C. Bishop

Dot product is product of:
(i) projection of x onto w (i.e., ||x|| cos θ), and (ii) the length of w
Dot product is 0 if x is perpendicular to w
Add b, if >0 then positive class, if <0 then negative class

Perceptron

Model: $f(x) = \sum_{i=1}^{m} w_i x_i + b$

$y = sign[f(x)]$

Learning: $if \;\; y(j)(\sum_{i=1}^{m} w_i x_i(j) + b) \leq 0$

$then \;\; w \leftarrow w + \eta y(j) x(j)$

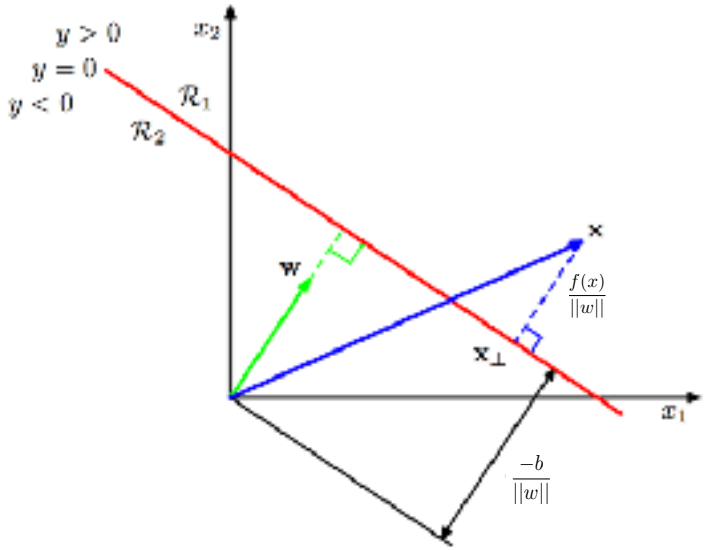


Figure: C. Bishop

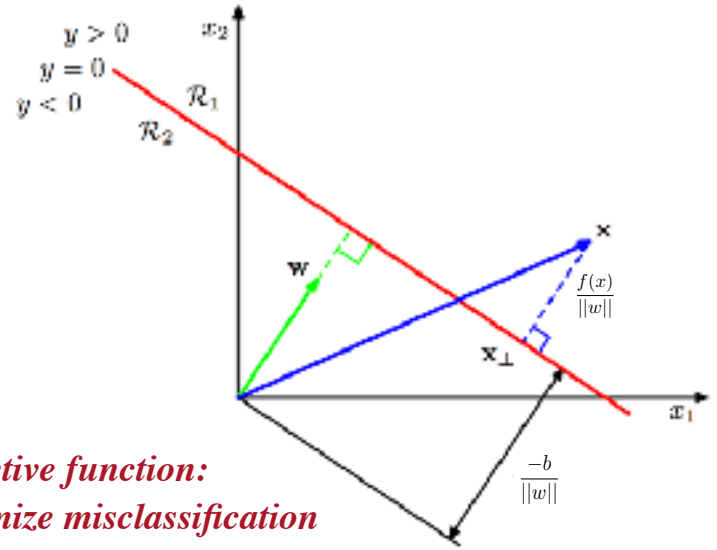
Perceptron

Model: $f(x) = \sum_{i=1}^{m} w_i x_i + b$

$y = sign[f(x)]$

Learning: $if \;\; y(j)(\sum_{i=1}^{m} w_i x_i(j) + b) \leq 0$

$then \;\; w \leftarrow w + \eta y(j) x(j)$



Objective function: minimize misclassification

Figure: C. Bishop

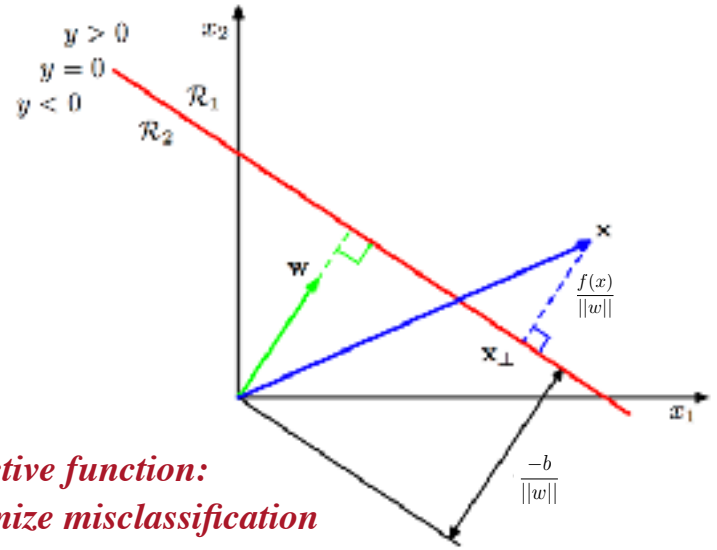
Perceptron

Model: $f(x) = \sum_{i=1}^{m} w_i x_i + b$

$y = sign[f(x)]$

Learning: $if \;\; y(j)(\sum_{i=1}^{m} w_i x_i(j) + b) \leq 0$

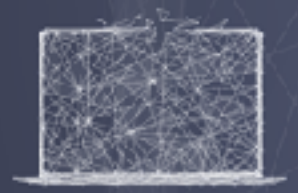
$then \;\; w \leftarrow w + \eta y(j) x(j)$



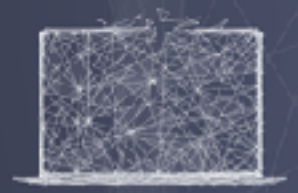
Objective function: minimize misclassification

Figure: C. Bishop

Optimization: Iterate over training examples for fixed number of iterations or until error is below a pre-specified threshold



Example: Naive Bayes classifier



Naive Bayes classifier

$$P(C|\mathbf{X}) \propto P(\mathbf{X}|C)P(C)$$

$$\propto \prod_{i=1}^{m} P(X_i|C)P(C)$$

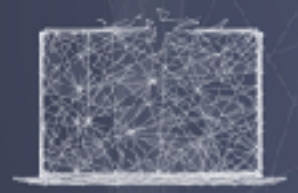
Assumption: Attributes are conditionally independent given the class

Naive Bayes classifier

$$P(C|\mathbf{X}) \propto P(\mathbf{X}|C)P(C)$$ Bayes rule

$$\propto \prod_{i=1}^{m} P(X_i|C)P(C)$$ Naive assumption

Assumption: Attributes are conditionally independent given the class

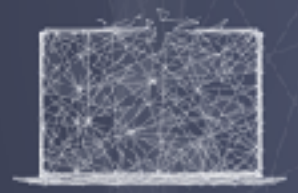


Objective function: Likelihood

- NBC likelihood uses the NBC probabilities for each data instance (i.e., probability of the class given the attributes)

$$
\begin{aligned}
L(\theta|D) &= \prod_{i=1}^{n} p(i|\theta) \\
&\propto \prod_{i=1}^{n} p(\mathbf{x}_i|c_i,\theta)P(c_i|\theta) \\
&\propto \prod_{i=1}^{n}\prod_{j=1}^{m} p(x_{ij}|c_i,\theta)P(c_i|\theta)
\end{aligned}
$$

- Optimization: “Learn” the best parameters by finding the values of θ that maximizes likelihood of the data



Objective function: Likelihood

- NBC likelihood uses the NBC probabilities for each data instance (i.e., probability of the class given the attributes)

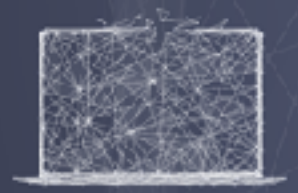
$$L(\theta|D) = \prod_{i=1}^{n} p(i|\theta)$$

General likelihood

$$\propto \prod_{i=1}^{n} p(\mathbf{x}_i|c_i,\theta)P(c_i|\theta)$$

$$\propto \prod_{i=1}^{n}\prod_{j=1}^{m} p(x_{ij}|c_i,\theta)P(c_i|\theta)$$

- Optimization: “Learn” the best parameters by finding the values of θ that maximizes likelihood of the data



Objective function: Likelihood

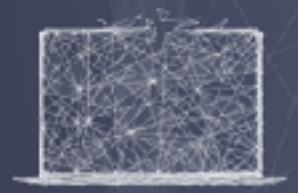
- NBC likelihood uses the NBC probabilities for each data instance (i.e., probability of the class given the attributes)

$$L(\theta|D) = \prod_{i=1}^{n} p(i|\theta) \qquad \textbf{General likelihood}$$

$$\propto \prod_{i=1}^{n} p(\mathbf{x}_i|c_i,\theta)P(c_i|\theta) \qquad \textbf{Bayes rule}$$

$$\propto \prod_{i=1}^{n}\prod_{j=1}^{m} p(x_{ij}|c_i,\theta)P(c_i|\theta)$$

- Optimization: “Learn” the best parameters by finding the values of θ that maximizes likelihood of the data



Objective function: Likelihood

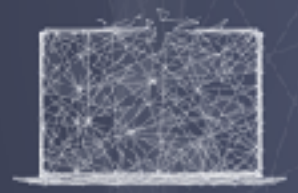
- NBC likelihood uses the NBC probabilities for each data instance (i.e., probability of the class given the attributes)

$$L(\theta|D) = \prod_{i=1}^{n} p(i|\theta) \qquad \textbf{General likelihood}$$

$$\propto \prod_{i=1}^{n} p(\mathbf{x}_i|c_i,\theta)P(c_i|\theta) \qquad \textbf{Bayes rule}$$

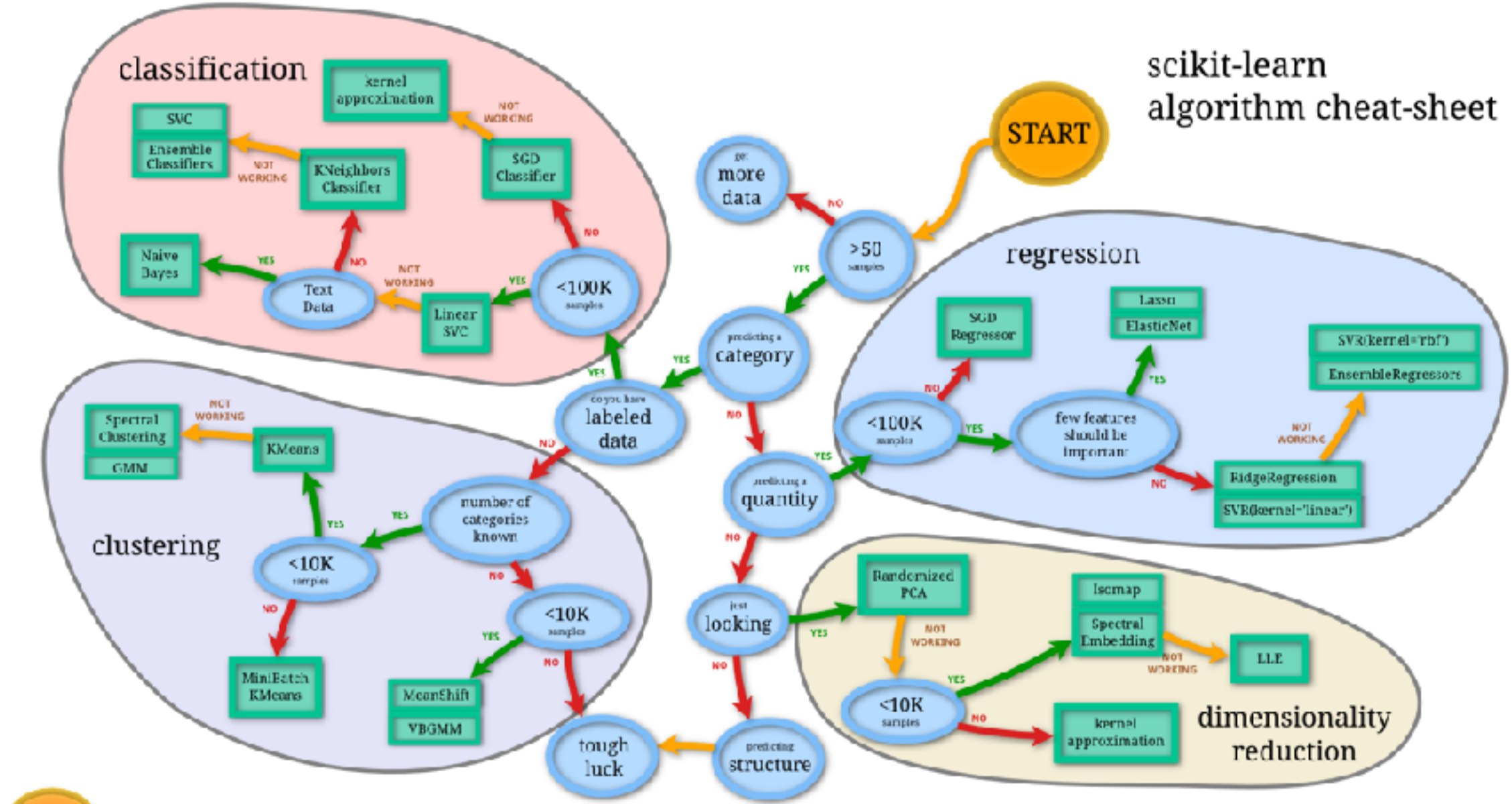
$$\propto \prod_{i=1}^{n}\prod_{j=1}^{m} p(x_{ij}|c_i,\theta)P(c_i|\theta) \qquad \textbf{Naive assumption}$$

- Optimization: “Learn” the best parameters by finding the values of θ that maximizes likelihood of the data



Machine learning methods

- Differ in their choice of:
  - Model representation
  - Objective function
  - Search/optimization method
- How to pick the right method for your task?
  - Think about data characteristics, data size, analysis task, modeling goals, computational resources, etc.



Source: https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html