Generative and Discriminative Learning

Machine Learning

THE UNIVERSITY OF UTAH

What we saw most of the semester

- A fixed, unknown distribution D over $X \to Y$
  - X: Instance space, Y: label space (eg: {+1, -1})

What we saw most of the semester

- A fixed, unknown distribution D over $X \rightarrow Y$
  - X: Instance space, Y: label space (eg: {+1, -1})

Is this different from assuming a distribution over X and a fixed oracle function f?

What we saw most of the semester

- A fixed, unknown distribution D over $X \rightarrow Y$
  - X: Instance space, Y: label space (eg: {+1, -1})

- Given a dataset S = $\{(x_i, y_i)\}$

Is this different from assuming a distribution over X and a fixed oracle function f?

What we saw most of the semester

- A fixed, unknown distribution D over $X \to Y$
  - X: Instance space, Y: label space (eg: {+1, -1})
- Given a dataset S = {(x_i, y_i)}
- Learning
  - Identify a hypothesis space H, define a loss function L(h, x, y)
  - Minimize average loss over training data (plus regularization)

Is this different from assuming a distribution over X and a fixed oracle function f?

What we saw most of the semester

- A fixed, unknown distribution D over $X \rightarrow Y$
  - X: Instance space, Y: label space (eg: {+1, -1})

- Given a dataset S = {(x_i, y_i)}

- Learning
  - Identify a hypothesis space H, define a loss function L(h, x, y)
  - Minimize average loss over training data (plus regularization)

- The guarantee
  - If we find an algorithm that minimizes loss on the observed data
  - Then, learning theory guarantees good future behavior (as a function of H)

Is this different from assuming a distribution over X and a fixed oracle function f?

Discriminative models

Goal: learn directly how to make predictions

- Look at many (positive/negative) examples
- Discover regularities in the data
- Use these to construct a prediction policy
- Assumptions come in the form of the hypothesis class

Bottom line: approximating $h: X \rightarrow Y$ is estimating the *conditional* probability $P(Y|X)$

Generative models

- Explicitly model how instances in each category are generated by modeling the *joint* probability of X and Y, that is $P(Y, X)$

- That is, learn $P(X|Y)$ and $P(Y)$

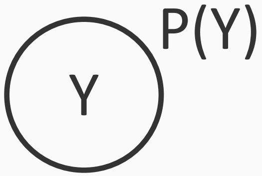
- The naïve Bayes classifier does this
  - Naïve Bayes is a generative model

- Predict $P(Y|X)$ using the Bayes rule

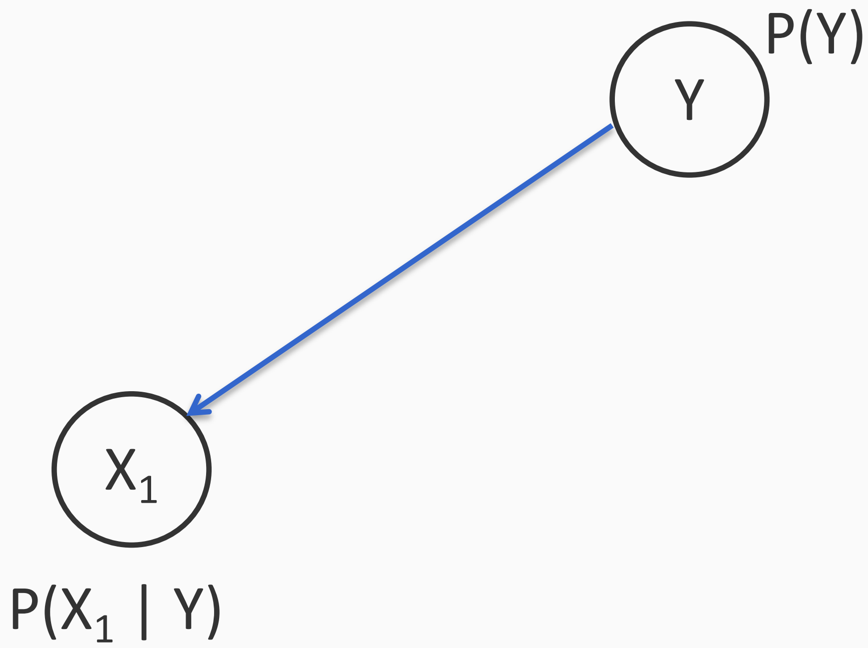
Example: Generative story of naïve Bayes

Example: Generative story of naïve Bayes

First sample a label

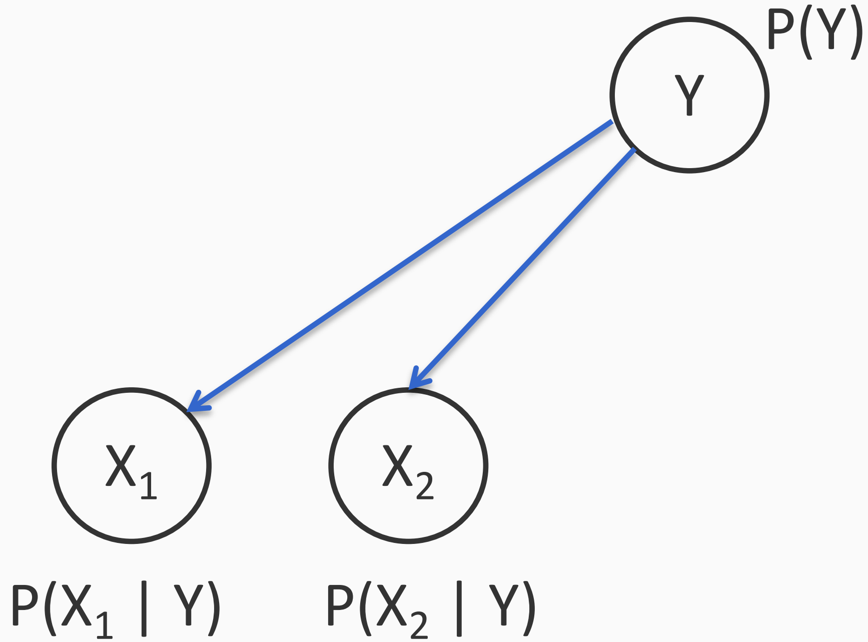


Example: Generative story of naïve Bayes



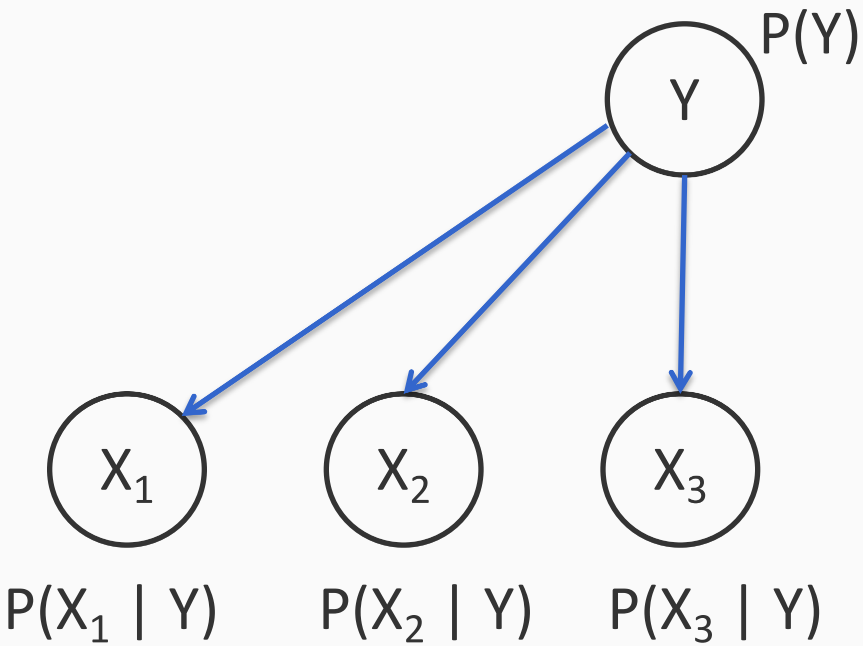
Given the label, sample the features independently from the conditional distributions

Example: Generative story of naïve Bayes



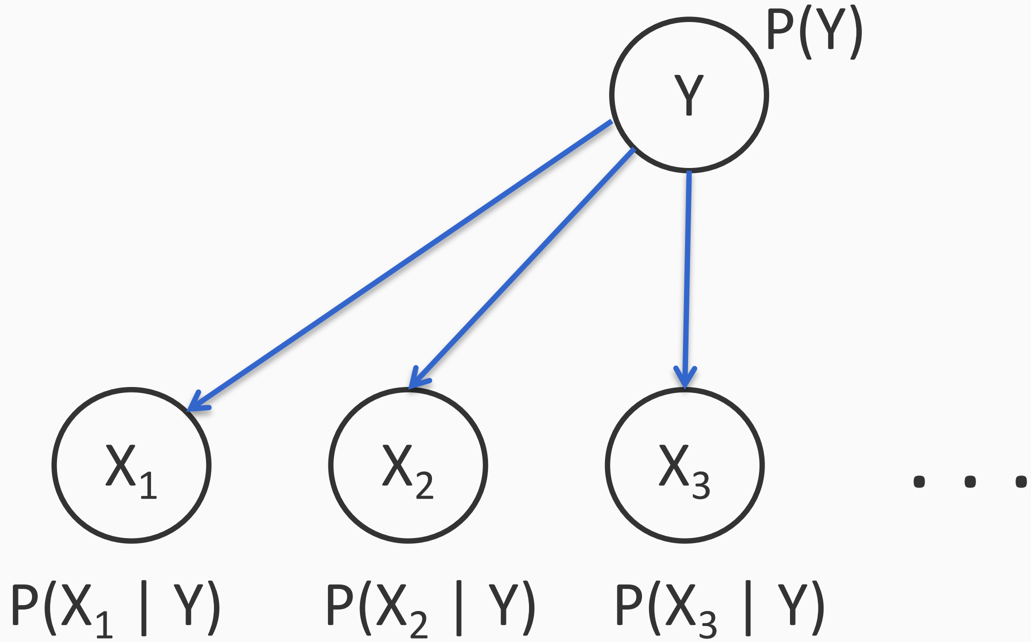
Given the label, sample the features independently from the conditional distributions

Example: Generative story of naïve Bayes



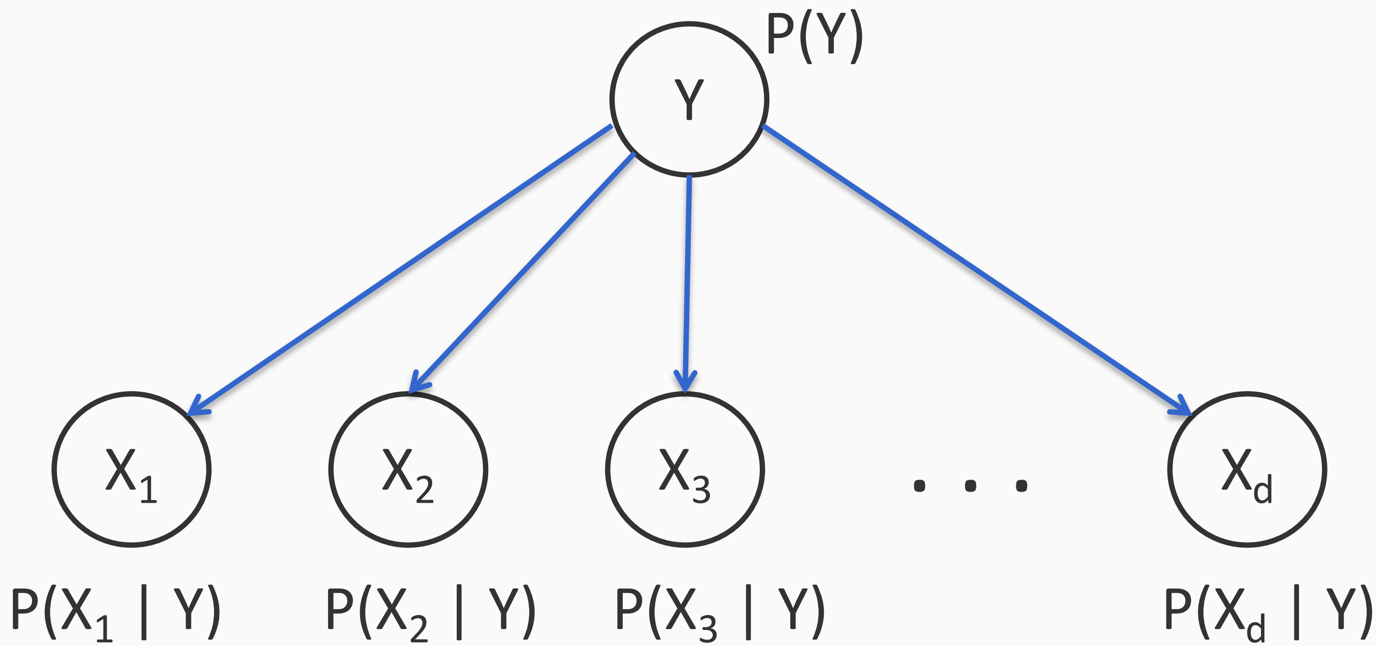
Given the label, sample the features independently from the conditional distributions

Example: Generative story of naïve Bayes



Given the label, sample the features independently from the conditional distributions

Example: Generative story of naïve Bayes



Given the label, sample the features independently from the conditional distributions

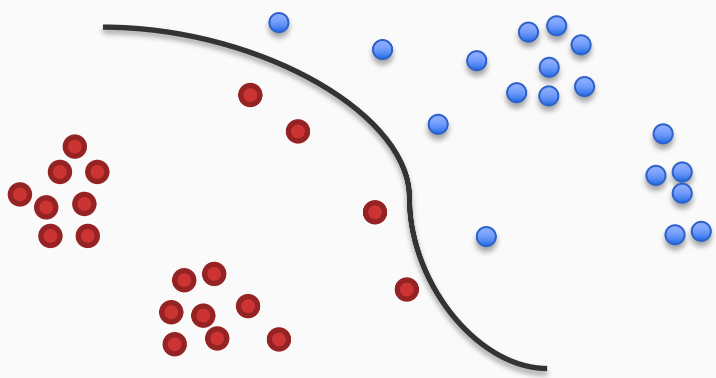
Generative vs Discriminative models

- Generative models
  - **learn P(x, y)**
  - Use the capacity of the model to characterize how the data is generated (both inputs and outputs)
  - Eg: Naïve Bayes, Hidden Markov Model

- Discriminative models
  - **learn P(y | x)**
  - Use model capacity to characterize the decision boundary only
  - Eg: Logistic Regression, Conditional models (several names), most neural models

Generative vs Discriminative models

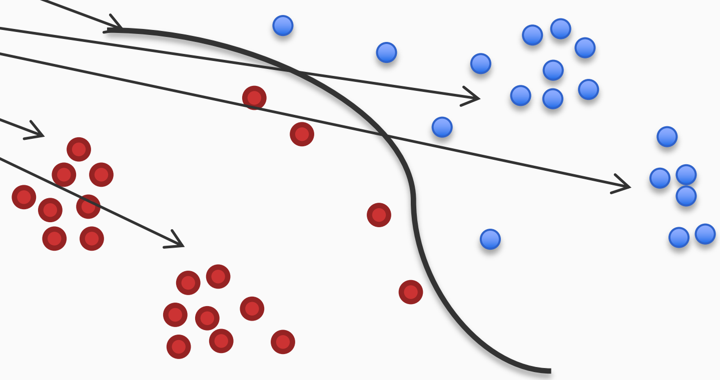
- Generative models
  - **learn P(x, y)**
  - Use the capacity of the model to characterize how the data is generated (both inputs and outputs)
  - Eg: Naïve Bayes, Hidden Markov Model



- Discriminative models
  - **learn P(y | x)**
  - Use model capacity to characterize the decision boundary only
  - Eg: Logistic Regression, Conditional models (several names), most neural models

Generative vs Discriminative models

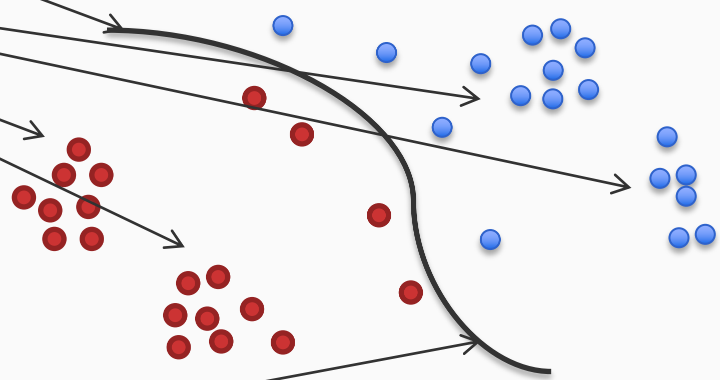
- Generative models
  - **learn P(x, y)**
  - Use the capacity of the model to characterize how the data is generated (both inputs and outputs)
  - Eg: Naïve Bayes, Hidden Markov Model



- Discriminative models
  - **learn P(y | x)**
  - Use model capacity to characterize the decision boundary only
  - Eg: Logistic Regression, Conditional models (several names), most neural models

Generative vs Discriminative models

- Generative models
  - **learn P(x, y)**
  - Use the capacity of the model to characterize how the data is generated (both inputs and outputs)
  - Eg: Naïve Bayes, Hidden Markov Model

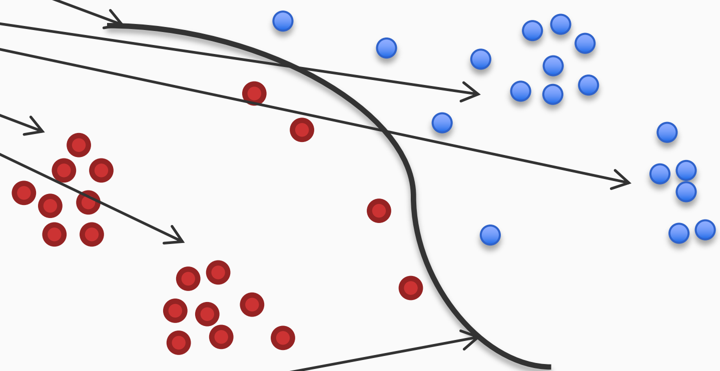


- Discriminative models
  - **learn P(y | x)**
  - Use model capacity to characterize the decision boundary only
  - Eg: Logistic Regression, Conditional models (several names), most neural models

Generative vs Discriminative models

- Generative models
  - **learn P(x, y)**
  - Use the capacity of the model to characterize how the data is generated (both inputs and outputs)
  - Eg: Naïve Bayes, Hidden Markov Model

A generative model tries to characterize the distribution of the inputs, a discriminative model doesn't care



- Discriminative models
  - **learn P(y | x)**
  - Use model capacity to characterize the decision boundary only
  - Eg: Logistic Regression, Conditional models (several names), most neural models