Discriminative and Generative ML models

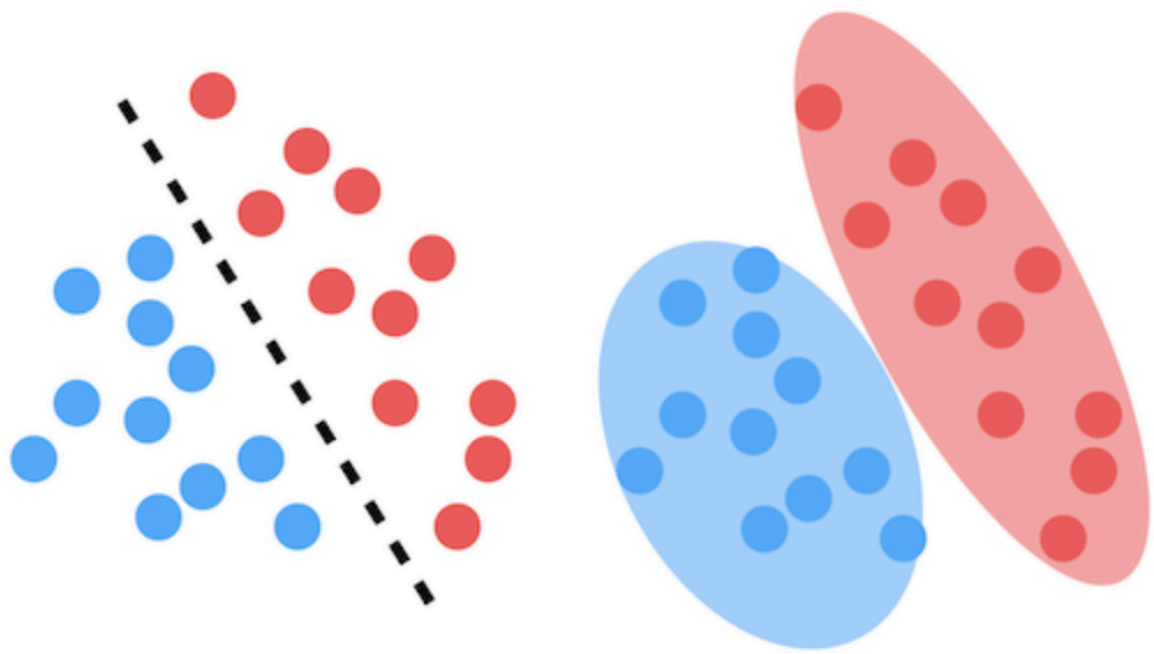
ML 101

Yash Saraswat

Basic Idea

Given inputs x and outputs y, there are two ways to model probability P(y|x).

1. Don't model x, model P(y|x) directly using a probability distribution. [Discriminative]
2. Model both x and y via their joint distribution P(x, y) and get the conditional P(y|x). [Generative]



Discriminative vs Generative

Img source: betterprogramming.pub

Examples

Discriminative Approach: Directly model P(y|x) using some distribution.

- Prob. Linear Regression $p(y|\boldsymbol{x}, \boldsymbol{w}) = \mathcal{N}(y|\boldsymbol{w}^\top \boldsymbol{x}, \beta^{-1})$
- Prob. Logistic Regression $p(y|\boldsymbol{x}, \boldsymbol{w}) = \text{Bernoulli}(y|\sigma(\boldsymbol{w}^\top \boldsymbol{x}))$

Generative Approach: Model P(y|x) using their joint distribution.

- $$p(y|\boldsymbol{x}, w) = \frac{p(\boldsymbol{x}, y|w)}{p(\boldsymbol{x}|w)}$$

$$p(y = k|\boldsymbol{x}, w) = \frac{p(\boldsymbol{x}, y=k|w)}{p(\boldsymbol{x}|w)} = \frac{p(\boldsymbol{x}|y = k, w)p(y=k|w)}{\sum_{\ell=1}^{K} p(\boldsymbol{x}|y = \ell, w)p(y=\ell|w)}$$

Discriminative Models

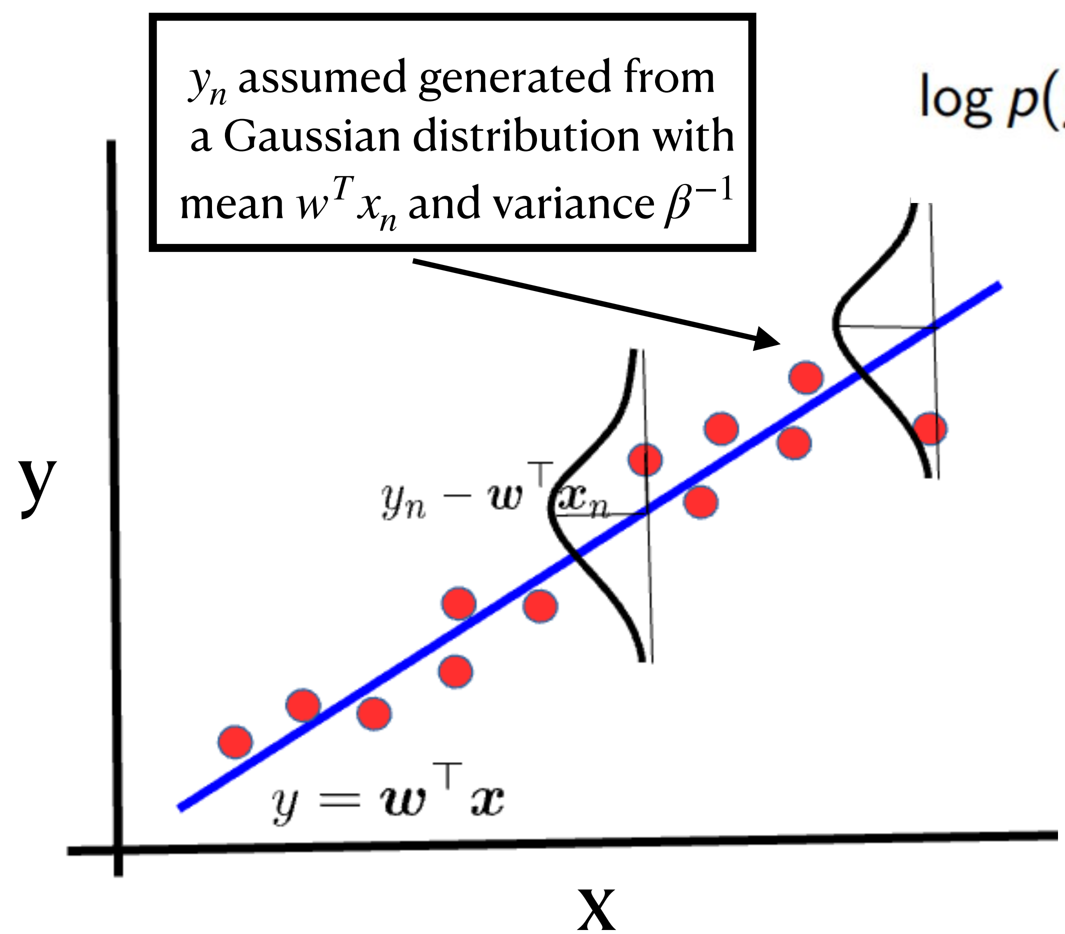
- Conditional models used for classification or regression.
- Distinguish decision boundaries based on observed data $P(y|x)$.
- Does not model the class distribution $P(x|y=k)$.
- Does not make any assumptions about the data points.
- Flow of training Discriminative models
  1. Assume some functional form for the probability, such as $P(y|x, w)$.
  2. With the help of training data, we estimate the parameters of $P(y|x, w)$.

Discriminative Models - 2

Prob. Linear Regression

- Given input examples x $\in R^d$ and output labels y $\in R^d$.

$$p(y_n|\boldsymbol{w},\boldsymbol{x}_n) = \mathcal{N}(y_n|\boldsymbol{w}^\top\boldsymbol{x}_n,\beta^{-1}) = \sqrt{\frac{\beta}{2\pi}}\exp\left[-\frac{\beta}{2}(y_n - \boldsymbol{w}^\top\boldsymbol{x}_n)^2\right]$$



$$\log p(\boldsymbol{y}|\mathbf{X},\boldsymbol{w}) \propto -\frac{\beta}{2}\sum_{n=1}^{N}(y_n - \boldsymbol{w}^\top\boldsymbol{x}_n)^2$$

Log Likelihood (ignoring constants wrt w)

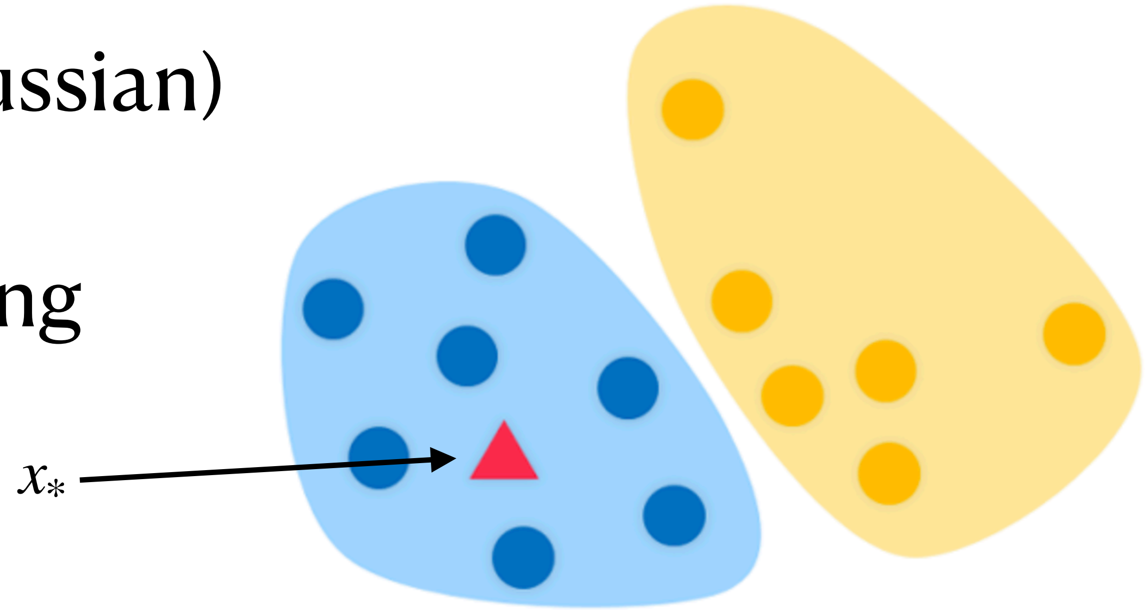
Discriminative Models-3

Examples

- Linear Regression
- Logistic Regression
- Support Vector Machines (SVM)
- Traditional Neural Networks
- Nearest Neighbor models
- Decision Trees and Random Forest

Generative Models

- Generative models learns the probability distribution of inputs from each class $P(x|y=k)$ [class-conditional]
- Learning class conditionals can help generate new data instances.
- Usually, assume some form (e.g. Gaussian) for the inputs and then estimate the parameters of input distribution using MLE/MAP/Bayesian Inference.
- Predict label of new input x_* by comparing its probability under each class.



Generative classification for 2 classes

Img source: medium.com

Generative Models-2

Classification example

- Consider a classification problem with $K \geq 2$ classes.
- Let the class prior probability P(y=k) for $k \in \{1,2,..,K\}$ comes from Multinomial distribution.

$$p(y \,|\, \pi) = multinoulli(y \,|\, \pi_1, \pi_2, \ldots, \pi_k) = \prod_{k=1}^{K} \pi_k^{I[y=k]}$$

- Let the class conditional probability $p(x \,|\, y = k, \theta)$ be a d-dimensional gaussian distribution for d-dim x.

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}_k|}} \exp[-(\boldsymbol{x} - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_k)]$$

Generative Models-3

Classification Example

- Prediction rule for Generative classification:

$$p(y = k|\boldsymbol{x}, \theta) = \frac{\pi_k |\boldsymbol{\Sigma_k}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma_k}^{-1}(\mathbf{x} - \boldsymbol{\mu}_k)\right]}{\sum_{k=1}^{K} \pi_k |\boldsymbol{\Sigma_k}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma_k}^{-1}(\mathbf{x} - \boldsymbol{\mu}_k)\right]}$$

- Point estimate prediction for new test input using MLE/MAP:

$$p(y_* = k|\boldsymbol{x}_*, \theta) = \frac{\pi_k |\boldsymbol{\Sigma_k}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x}_* - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma_k}^{-1}(\mathbf{x}_* - \boldsymbol{\mu}_k)\right]}{\sum_{k=1}^{K} \pi_k |\boldsymbol{\Sigma_k}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x}_* - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma_k}^{-1}(\mathbf{x}_* - \boldsymbol{\mu}_k)\right]}$$

Generative Models-4

Examples

- Generative Classification
- Naive Bayes
- Latent Dirichlet Allocation (LDA)
- Bayesian Networks
- Variational Auto Encoders (VAE)
- Generative Adversarial Networks (GAN)
- AutoRegressive Model

Discriminative vs Generative

	Discriminative	**Generative**
Approach	Directly model P(y|x) using some appropriate dist.	Model P(y|x) using their joint distribution P(x,y).
Number of Parameters	Few parameters. Need to learn just weight w vectors	More parameters. Parameters for class conditionals, class priors.
Ease of parameter Estimation	Generally use iterative methods.	For simple class-cond., easy to compute closed form solution
Dealing with Missing Features	Cannot handle missing features.	Can handle missing features by integrating out missing features.

Discriminative vs Generative

	Discriminative	Generative
Inputs having features with mixed types	Difficult to handle because we model P(y|x) directly.	Can handle by using appropriate $p(x_d \| y)$ for each type of input.
Leveraging Unlabeled data	Difficult to handle	Can handle easily be considering missing labels as latent variables.
Adding data from new class	Model needs to be retrained on new dataset.	Model just require estimating class conditional for new class.

Pro-Tip

In tabular data, data imputation for missing feature values can be done using Generative models. Finding out Class Conditional can help in filling missing features values accurately.