CS456: Machine Learning

Discriminant Analysis

Jakramate Bootkrajang

Department of Computer Science
Chiang Mai University

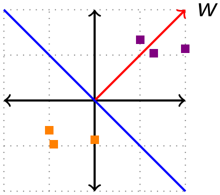
Objectives

- To understand the philosophy behind discriminant analysis classifier
- To understand how to learn discriminant analysis model
- To understand Bayes rule
- To understand the difference between discriminative and generative classfiers

Outlines

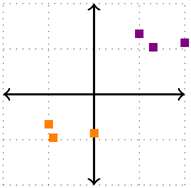
- Discriminant Analysis
- Parameters estimation
- Bayes rule
- Plotting decision boundary
- Generative vs Discriminative

Probability supporting prediction



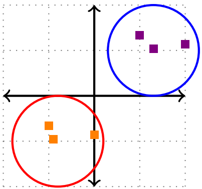
- Recall how LR converts the distance into probability which increases as the distance from decision boundary increases
- using the sigmoid function $p(y=1|\mathbf{x},\mathbf{w}) = \frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}}$

Discriminant Analysis approach (1/2)



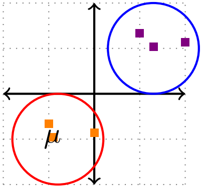
- A discriminant analysis model takes different approach
- It does not look for a weight vector $\mathbf{w}$ and therefore there is no decision boundary to begin with

Discriminant Analysis approach (2/2)



- Instead, DA tries to model data distributions for each classes
- Usually, it assumes that data is approximately normally distributed
- This corresponds to modelling step

(Multivariate) Normal Distribution



- Each of the circles, represent a single multivariate normal distribution with μ as its center and with Σ specifying data spread
- $p(\mathbf{x}|\mu, \Sigma) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right)}{\sqrt{(2\pi)^m|\Sigma|}}$ is its density function [1]

[1] gives (relative) probability that $\mathbf{x}$ comes from this distribution (aka data likelihood)

How to estimate μ and Σ (model learning)

- Let N_+ be the size of data in positive class and N_- be the size of data in negative class, we estimate
- $\mu_+ = \frac{1}{N_+}\sum_{i=1}^{N_+}\mathbf{x}_i$
- $\mu_- = \frac{1}{N_-}\sum_{i=1}^{N_-}\mathbf{x}_i$
- $\Sigma_+ = \frac{1}{N_+}\sum_{i=1}^{N_+}(\mathbf{x}_i - \mu_+)(\mathbf{x}_i - \mu_+)^T$
- $\Sigma_- = \frac{1}{N_-}\sum_{i=1}^{N_-}(\mathbf{x}_i - \mu_-)(\mathbf{x}_i - \mu_-)^T$

How to get probability of class label ?

- What we have so far is the relative probiliity $p(\mathbf{x}|\mu, \Sigma)$ (data likelihood)
- But to predict if class label is 1 we need $p(y = 1|\mathbf{x})$
- Similarly, to predict if class label is -1 we need $p(y = -1|\mathbf{x})$
- How to get such quantity ?

The Bayes rule

There is Bayes rule which links the two quantities

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \tag{1}$$

- $p(y|x)$ is class posterior probability
- $p(x|y)$ is data likelihood
- $p(y)$ is class prior probability
- $p(x)$ is data evidence: normaliser to make $p(y|x) \in [0, 1]$

The Bayes rule in our context

Using Bayes rule

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

in our context,

$$p(y=1|\mathbf{x}) = \frac{p(\mathbf{x}|y=1)p(y=1)}{p(\mathbf{x})} \tag{2}$$

$$= \frac{p(\mathbf{x}|\mu_+, \Sigma_+)p(y=1)}{p(\mathbf{x}|y=1)p(y=1)p(\mathbf{x}|y=-1)p(y=-1)} \tag{3}$$

Class probability for negative class

Also using Bayes rule we can calculate probability that $\mathbf{x}$ belongs to negative class

$$p(y=-1|\mathbf{x}) = \frac{p(\mathbf{x}|y=-1)p(y=-1)}{p(\mathbf{x})} \tag{4}$$

$$= \frac{p(\mathbf{x}|\mu_-, \Sigma_-)p(y=-1)}{p(\mathbf{x}|y=1)p(y=1)p(\mathbf{x}|y=-1)p(y=-1)} \tag{5}$$

How to compute $p(y=1)$ and $p(y=-1)$?

- $p(y=1)$ and $p(y=-1)$ are called class prior probability
- It specifies how likely we observe positive class (and negative class) in general
- Can be inferred from data $p(y=1) = \frac{N_+}{N}$, $p(y=-1) = \frac{N_-}{N}$
- Or assumed based on prior knowledge
  - If $y=1$ is class of female, and $y=-1$ is class of male, we expect $p(y=1) = 0.5$ and $p(y=-1) = 0.5$

Discriminant Analysis in summary

1. Input data in the form $\{(\mathbf{x}_i, y_i)\}, i = 1 : N$ and $y_i \in \{-1, 1\}$
2. Infer $p(y = 1)$ and $p(y = -1)$ from data
3. Estimate μ_+, Σ_+, μ_- and Σ_- from data
4. To predict class label of $\mathbf{x}_q$

$$p(y = 1|\mathbf{x}_q) = \frac{p(\mathbf{x}_q|\mu_+, \Sigma_+)p(y = 1)}{p(\mathbf{x}_q|y = 1)p(y = 1)p(\mathbf{x}_q|y = -1)p(y = -1)}$$

$$p(y = -1|\mathbf{x}_q) = \frac{p(\mathbf{x}_q|\mu_-, \Sigma_-)p(y = -1)}{p(\mathbf{x}_q|y = 1)p(y = 1)p(\mathbf{x}_q|y = -1)p(y = -1)}$$

5. Predict $y_q = 1$ if $p(y = 1|\mathbf{x}_q) > p(y = -1|\mathbf{x}_q)$
6. Else predict $y_q = -1$

Wait, where's the decision boundary?

- You probably noticed that there's no *explicit* decision boundary for DA
- For, DA we did not assume any kind of functional decision boundary

Plotting the (implicit) decision boundary

- However, we can still visualise the boundary by observing the points where prediction change from -1 to 1
- `https://colab.research.google.com/drive/1Eacymr2skXvIv0N2y-a2NQSK5kokJhkB`

Shape of the decision boundary

- The shape of decision boundary changes with types of the covariances

1. Two classes use common covariance (linear)
   - aka. linear discriminant analysis
2. Two classes use their own covariances (non-linear)

Generative VS discriminative

- There are two types of parametric classification model

1. Discriminative classifier
    - Model the decision boundary explicitly
    - Example: SVM, Logistic regression
2. Generative classifier
    - Model the occurrence of data using probability distribution
    - decision boundary is not explicit
    - Example: Discriminant analysis

Objectives: revisited

- To understand the philosophy behind discriminant analysis classifier
- To understand how to learn discriminant analysis model
- To understand Bayes rule
- To understand the difference between discriminative and generative classfiers

Lastly

Questions please ..