Tut 5: LDA (Bayes' Classifier)

Jan 2022

The general mathematical formulation of a generative model:

$$\begin{aligned} h_D(\boldsymbol{x}) &= \underset{j\in\{1,\ldots,K\}}{\operatorname{argmax}}\ \mathbb{P}(Y=j|\boldsymbol{X}=\boldsymbol{x}) = \underset{j\in\{1,\ldots,K\}}{\operatorname{argmax}}\ \mathbb{P}(\boldsymbol{X}=\boldsymbol{x}|Y=j)\mathbb{P}(Y=j) \\ &= \underset{j\in\{1,\ldots,K\}}{\operatorname{argmax}}\ [\ln \mathbb{P}(\boldsymbol{X}=\boldsymbol{x}|Y=j) + \ln \mathbb{P}(Y=j)] \end{aligned} \tag{6.1}$$

QDA (only works for numeric inputs which follows the normal distribution):

$$\boxed{\mathbb{P}(\boldsymbol{X}=\boldsymbol{x}|Y=j) \approx \frac{1}{(2\pi)^{p/2}\sqrt{|\boldsymbol{C}_j|}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_j)^T \boldsymbol{C}_j^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_j)\right\}.}$$

LDA (only works for numeric inputs which follows the normal distribution):

$$\boxed{\mathbb{P}(\boldsymbol{X}=\boldsymbol{x}|Y=j) \approx \frac{1}{(2\pi)^{p/2}\sqrt{|\boldsymbol{C}|}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_j)^T \boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_j)\right\}.}$$

$$\Rightarrow \boxed{h_D(\boldsymbol{x}) = \underset{j\in\{1,\ldots,K\}}{\operatorname{argmax}} \left\{\ln \mathbb{P}(Y=j) + \vec{\mu}_j^T \boldsymbol{C}^{-1}\left[\boldsymbol{x} - \frac{1}{2}\vec{\mu}_j\right]\right\}.}$$

Naive Bayes:

$$\mathbb{P}(\boldsymbol{X}=\boldsymbol{x}|Y=j) \approx \prod_{i=1}^{p} \mathbb{P}(X_i = x_i|Y=j)$$

1. Factory XYZ produces very expensive and high quality golf balls that their qualities are measured in term of curvature and diameter. Result of quality control by experts is given in the table below:

Curvature	Diameter	Result
2.95	6.63	Passed
2.53	7.79	Passed
3.57	5.65	Passed
3.16	5.47	Passed
2.58	4.46	Not Passed
2.16	6.22	Not Passed
3.27	3.52	Not Passed

As a consultant to the factory, you get a task to set up the criteria for automatic quality control using LDA model. Then, the manager of the factory also wants to test your criteria upon a new type of golf ball which have curvature 2.81 and diameter 5.46.

(a) Plot the data with axes of curvature and diameter. Comment on the plot.

(b) Write the data into matrix form by separating into "Passed" and "Not Passed".

(c) Compute the prior probability for both classes.

(d) Compute the mean vectors for both classes.

(e) Compute the group covariance matrix.

(f) Write down the discriminant functions for both classes.

(g) Transform all the given data into discriminant functions.

(h) Locate the new golf ball in the plot as well as the functions to classify it.

(i) Plot the discriminant line into plot of $\delta_1(X)$ versus $\delta_2(X)$.