Linear Discriminant Analysis (1)

Son Nguyen

Classification Problem

- Given a dataset that has x and y (class or label)

x	y
1.49	1
1.23	1
0.95	1
0.89	1
2.58	2
2.65	2
2.27	2
1.94	2
1.39	2
1.44	2

$y = 1$

$y = 2$

- Given a new value x, what class the it belongs to? (what is the predicted y value)
- If $x = 1.4$, what is its associated y value? (What class it belongs to?)

Approach

- We will estimate two probabilities $p_1 = P(y = 1|x = 1.4)$ and $p_2 = P(y = 2|x = 1.4)$.
- If $p_1 > p_2$, we will classify the new point to class 1 and vice versa.

$(p_1 + p_2 = 1)$

Approach

We have, using the Bayes' Rule,

$$p_1 = P(y=1|x=1.4) = \frac{\overset{\text{prob.}}{P(y=1)} * \overset{\text{likelihood}}{L(x=1.4|y=1)}}{L(x=1.4)}$$

where L(A) denotes the likelihood of the event A. Similarly,

$$p_2 = P(y=2|x=1.4) = \frac{P(y=2) * L(x=1.4|y=\cancel{1}\,2)}{L(x=1.4)}$$

Since the denominator is the same we just need to compare the numerator.

$P(y=1) \approx \frac{4}{10} = .4$; $P(y=2) = \frac{6}{10} = .6$

$\underbrace{L(x=1.4|y=1)} = ?$

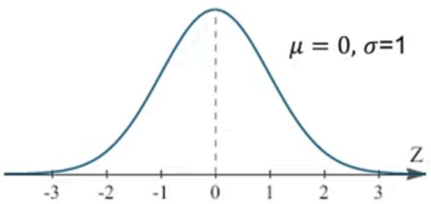
LDA Assumptions

- x is normally distributed in each class
- Assume that x has the same variance in both classes

Likelihood:

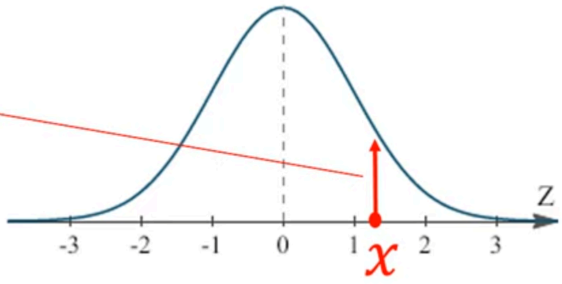
The *probability density function* for a normal distribution $N(\mu, \sigma^2)$ is:

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{\sigma^2}}$$



For a given distribution the *likelihood* of the distribution parameters being μ, σ^2 given the observation x is:

$$L(\mu,\sigma^2|x) = f(x|\mu,\sigma^2)$$



The estimated mean for X in class 1 is

$$\bar{x}_1 = \frac{1.49 + 1.23 + .95 + .89}{4} = 1.14$$

For class 2: $\bar{x}_2 = \frac{2.58 + 2.65 + 2.27 + 1.94 + 1.39 + 1.49}{6} = 2.04$

An estimated for the common s.d is the s.d of class 2:

$$\sigma(x_2) = .55$$

Therefor in class 1 $X \sim N(\mu = 1.14, \sigma = .55)$

class 2 $X \sim N(\mu = 2.04, \sigma = .55)$

$$L(x=1.4 \mid Y=1) = f(x=1.4 \mid \mu = 1.14,\ \sigma = .55)$$

$$= \frac{1}{\sqrt{2\pi \cdot .55^2}} \cdot e^{-(1.4-1.14)^2 / 2\cdot(.55)^2}$$

$$= .65$$

$$L(x=1.4 \mid Y=2) = f(x=1.4 \mid \mu = 2.04,\ \sigma = .55)$$

$$= \frac{1}{\sqrt{2\pi * .55^2}} \cdot e^{-(1.4-2.04)^2 / 2\cdot(.55)^2}$$

$$= .37$$

$$P_1 = \frac{P(Y=1) * L(X=1.4 \mid Y=1)}{A} = \frac{.4 * .65}{A} = \frac{.26}{A}$$

$$P_2 = \frac{P(Y=2) * L(X=1.4 \mid Y=2)}{A} = \frac{.6 * .37}{A} = \frac{.22}{A}$$

Because $P_1 > P_2 \Rightarrow$ The associated value for $x = 1.4$

is $y = 1$

* <u>Decision Boundary</u>

$P_1 = P_2$

$$\Rightarrow \frac{P(Y=1)\cdot f(X = x_0 \mid Y=1)}{A} = \frac{P(Y=2)\cdot f(X = x_0 \mid Y=2)}{A}$$

$$\Rightarrow .4 \cdot f(x = x_0 \mid \mu = 1.14, \sigma = .55) = .6\, f(x = x_0 \mid \mu = 2.04, \sigma = .55)$$

$$\Rightarrow .4 * \cancel{\frac{1}{\sqrt{2\pi\sigma^2}}} \cdot e^{-\frac{(x_0 - 1.14)^2}{2\cdot\sigma^2}} = .6\, \cancel{\frac{1}{\sqrt{2\pi\sigma^2}}} \cdot e^{-\frac{(x - 2.04)^2}{2\sigma^2}}$$

$$\ln\left[.4 \cdot e^{-\frac{(x_0-1.14)^2}{2\sigma^2}}\right] = \ln\left[.6 \; e^{-\frac{(x_0-2.04)^2}{2\sigma^2}}\right]$$

$$\Rightarrow \ln .4 - \frac{(x_0-1.14)^2}{2\sigma^2} = \ln(.6) - \frac{(x_0-2.04)^2}{2\sigma^2}$$

$$\Rightarrow -\frac{(x_0-1.04)^2}{2\sigma^2} + \frac{(x_0-2.04)^2}{2\sigma^2} = \ln .6 - \ln .4$$

$$\Rightarrow -(x_0-1.14)^2 + (x_0-2.04)^2 = .245$$

$$\Rightarrow (2x_0 - 3.18)(1.14 - 2.04) = .245$$

$$\Rightarrow \boxed{x_0 = 1.454}$$

Calculation

Calculation

https://planetcalc.com/4986/

https://www.standarddeviationcalculator.io/normal-distribution-calculator

Decision Boundary