# Discriminant Analysis

$P(Y=j \mid x)$

## Setup

Suppose $Y \in \{1, 2, \ldots, K\}$ falls into one of $K$ classes.

D.A. assumes a distribution on the feature given the current class, and inverts via Bayes rule.

Assume $[X \mid Y]$ invert to get $[Y \mid X]$

## Notation

- $\pi_k = P(Y=k)$ = the $\underbrace{\text{prior probability}}_{\text{class } k}$ that $Y$ is in class $k$
- $f_k(x) = P(X=x \mid Y=k)$ = pmf or pdf for $X$ given $Y$ is in class $k$.

Remark

$$p_k(x) = P(Y=k \mid X=x) = \frac{P(X=x \mid Y=k)\,P(Y=k)}{P(X=x)}$$

little k

$$= \frac{f_k(x)\,\pi_k}{\sum_{i=1}^{K} \pi_i f_i(x)} = \text{posterior probability that } Y \text{ is in class } k \text{ given } X=x.$$

total # of classes, big K

Remark For new feature $X=x$, decision rule is to classify $Y$ into the class that has highest posterior prob, i.e. the class for which $p_1(x), p_2(x), \dots, p_K(x)$ is biggest. The most probable class is that which maximizes $\pi_1 f_1(x), \pi_2 f_2(x), \dots, \pi_K f_K(x)$

## Linear discriminant analysis for $p=1$

one feature $X$

LDA assumes $[X \mid Y=k]$ are normal with a common variance across $k$.

$\mu_k$ depends on class

$$f_k(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left( -\frac{1}{2} \left( \frac{x-\mu_k}{\sigma} \right)^2 \right)$$

does not depend on $k$.

**Note** Classification of $Y$ involves finding the $k$ for which $\pi_1 f_1(x), \ldots, \pi_K f_K(x)$ is maximized

$$\log(\pi_k f_k(x)) = C_1 + \log\pi_k - \frac{1}{2\sigma^2}(x^2 - 2x\mu_k + \mu_k)^2$$

$$= C_1 + \log\pi_k + x\frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + C_2(x)$$

$$= \log\pi_k + x\frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \left(\text{stuff that doesn't depend on class } k\right)$$

DEF $\quad \delta_k(x) = \log\pi_k + x\frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2}$ → linear in x, hence LDA

is the <u>discriminant function</u> [for $k = 1, 2, \ldots, K$]

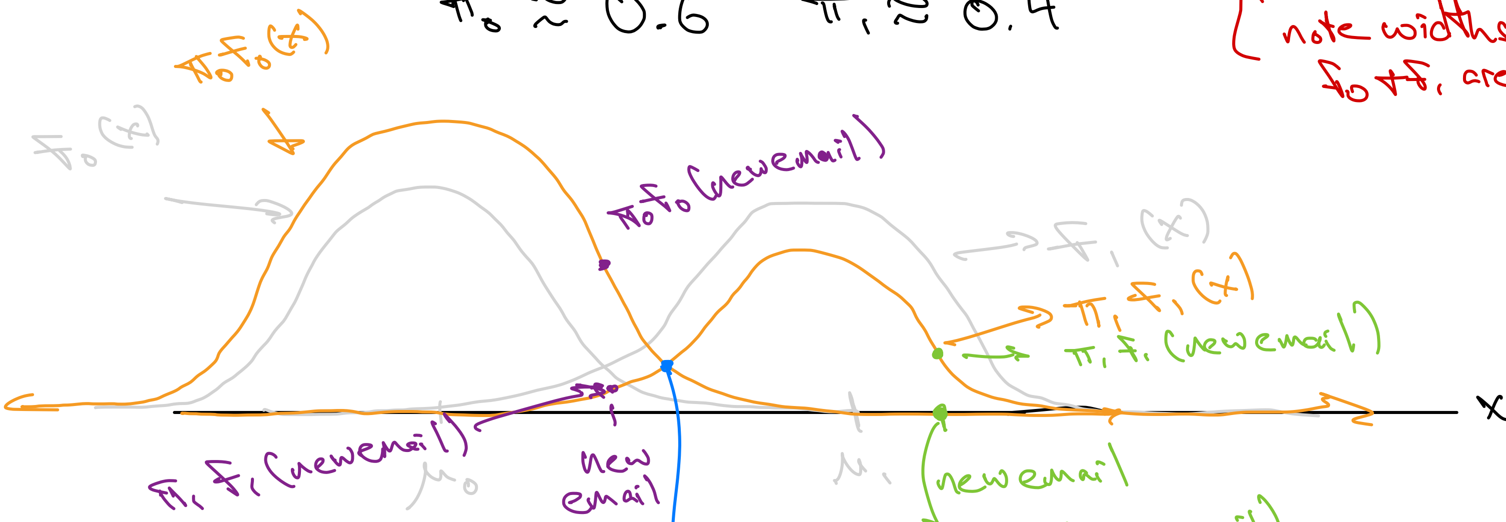
Note LDA categorizes $x$ into the class for which $\delta_1(x), \ldots, \delta_K(x)$ is maximized

$\boxed{\text{Ex}}$ $Y \in \{0,1\}$ = not spam / spam

$x = \log(\text{capLong}) = \log(\text{length of longest streak of capital letters})$

$\pi_0 \approx 0.6$ $\pi_1 \approx 0.4$

[note widths of $f_0$ & $f_1$ are same]




$\Rightarrow \hat{Y} = 0$

decision boundary

Compare $\pi_0 f_0(x)$ & $\pi_1 f_1(x)$ & classify into bigger of the two

$\hat{Y} = 1$

$\boxed{\text{Ex}}$ $K=2$, $\pi_1 = \pi_2$. Classify $Y$ into class 1 if

$$\delta_1(x) > \delta_2(x)$$

$$\log \pi_1 + x\frac{\mu_1}{\sigma^2} - \frac{\mu_1^2}{2\sigma^2} > \log \pi_2 + x\frac{\mu_2}{\sigma^2} - \frac{\mu_2^2}{2\sigma^2}$$

$$2x(\mu_1 - \mu_2) > \mu_1^2 - \mu_2^2$$

decision boundary is $\delta_1(x) = \delta_2(x)$

$$2x(\mu_1 - \mu_2) = \mu_1^2 - \mu_2^2$$

$$\Leftrightarrow \quad x = \frac{\mu_1 + \mu_2}{2}$$

## Estimation

- $\hat{\pi}_k = \frac{n_k}{n} = \frac{\#\text{ of Ys in class } k}{\text{~~total~~ }\#\text{ of Ys}}$

- $\hat{\mu}_k = \frac{1}{n_k} \sum_{i|y_i=k} x_i$ = avg of Xs for Y being in class k.

- $\hat{\sigma}^2 = \frac{1}{n-K} \sum_{k=1}^{K} \sum_{i|y_i=k} (x_i - \hat{\mu}_k)^2$

  = pooled empirical variance