QDA, LDA, MLE of parameters

Gaussian Discrimination Analysis

assumption: each class comes from normal dist

$X \sim N(\mu, \sigma^2)$ $\quad f(x) = \frac{1}{(\sqrt{2\pi}\sigma)^d} \exp\left\{-\frac{\|x-\mu\|^2}{2\sigma^2}\right\}$

d ← dimension

For each class C, suppose we estimate μ_c, σ_c^2 $\quad \pi_c = P(Y=c)$

given x, Bayes decision rule $r^*(x)$ predicts class C that maximizes

$f(X=x \mid Y=c)\,\pi_c$

$\ln w$ is monotonically increasing when $w > 0$, so it is equivalent to maximize

$$Q_c(x) = \ln\left((\sqrt{2\pi})^d f_c(x)\,\pi_c\right) = \frac{-\|x-\mu_c\|^2}{2\sigma^2} - d\ln\sigma_c + \ln\pi_c$$

Quadratic in x

$f_c(x)$ estimates $f(X=x \mid Y=c)$

Quadratic Discriminant Analysis (QDA)

2 classes C, D then:

$$r^*(x) = \begin{cases} C & Q_c(x) - Q_D(x) > 0 \\ D & \text{otherwise} \end{cases}$$

Decision fn quadratic in x. Bayes decision boundary is when

$Q_C(x) - Q_D(x) = 0$

To recover posterior prob. in 2-class case

$$P(Y=c|X) = \frac{f(X|Y=c)\pi_c}{f(X|Y=c)\pi_c + f(X|Y=D)\pi_D}$$

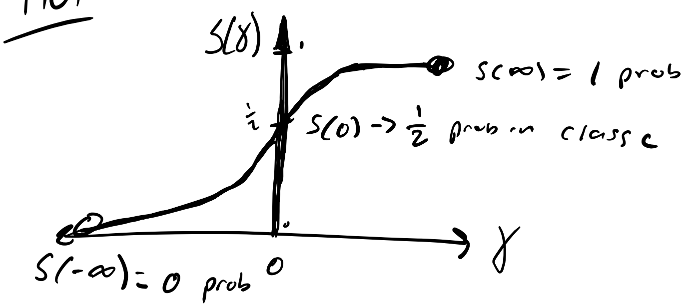
$* \; e^{Q_c(x)} = (\sqrt{2\pi})^d f_c(x)\,\pi_c$

$$= \frac{e^{Q_c(x)}}{e^{Q_c(x)} + e^{Q_D(x)}} = \frac{1}{1 + e^{Q_c(x) - Q_D(x)}}$$

$s(Q_c(x) - Q_D(x))$

$s(\gamma) = \frac{1}{1+e^{\gamma}}$ <= logistic/sigmoid function

Plot



Linear Discriminant Analysis (LDA)

- Less overfitting
- assumption: all gaussians have the same σ^2

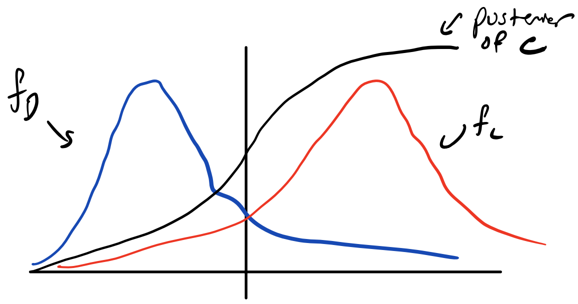
$$Q_c(x) - Q_D(x) = \underbrace{\frac{(\mu_c - \mu_D)^T x}{\sigma^2}}_{w^T x} + \underbrace{- \frac{\|\mu_c\|^2 - \|\mu_D\|^2}{2\sigma^2} + \ln \pi_c - \ln \pi_D}_{\alpha}$$

Choose c that maximizes linear Discriminant fn:

$$\frac{\mu_c^T x}{\sigma^2} - \frac{\|\mu_c\|^2}{2\sigma^2} + \ln \pi_c$$

2-class: decision boundary is $w^T x + \alpha = 0$

$$P(Y=c|X=x) = s(w^T x + \alpha)$$



Maximum Likelihood estimation of Parameters

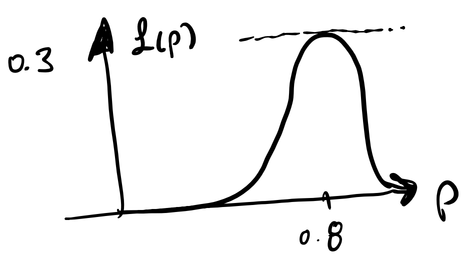
flip bias coin! heads w/ p 10 flips, 8 heads, 2 tails

most likely value of p?

$X \sim Bin(n, p)$ $\quad P[X=x] = \binom{n}{x}(1-p)^{n-x}p^{x}$

$P[X=8] = 45p^{8}(1-p)^{2} = \mathcal{L}(p)$ ← dist. parameter(s), likelihood fn

MLE: Method of estimating paramters of statistical dist. by maximizing $\mathcal{L}$ ... density estimation.



Solve by calc.

$$\frac{d\mathcal{L}}{dp} = 360(1-p)^{2}p^{7} - 90p^{8}(1-p) = 0$$

$$\Rightarrow p = 0.8$$

$$\pi_c = 0.8$$

Likelihood of a Gaussian:

$X_1, X_2, \dots X_n$ → best-fit Gaussian

Likelihood of generating these pts is $\mathcal{L}(\mu, \sigma; x_1 \dots x_n)$

$$= f(x_1)\, f(x_2) \cdots f(x_n)$$

↑ μ, σ to max whole thing

$\ell(\cdot) :=$ log likelihood $\ln(\mathcal{L}(\cdot))$

$$\ell(\mu, \sigma; x_1 \ldots x_n) = \ln f(x_1) + \ln f(x_2) \ldots \ln f(x_n)$$

$$= \sum_{i=1}^{n} \left(\frac{-\|x_i - \mu\|^2}{2\sigma^2} - d\ln\sqrt{2\pi} - d\ln\sigma \right)$$

Set $\nabla_\mu \ell = 0$, $\frac{\partial \ell}{\partial \sigma} = 0$

$$\nabla_\mu \ell = \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma^2} = 0 \Rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

sub $\hat{\mu}$

$$\frac{\partial \ell}{\partial \sigma} = \sum_{i=1}^{n} \frac{\|x_i - \mu\|^2 - d\sigma^2}{\sigma^3} = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{dn} \sum_{i=1}^{n} \|x_i - \mu\|^2$$

Use sample mean and variance to estimate gaussian

QDA:

- est. class conditional $\hat{\mu}_c$ and $\hat{\sigma}_c^2$
  - ↳ each class seperately
- est. priors $\hat{\pi}_c = \frac{n_c}{\sum_v n_v}$ ← all pts in all classes

LDA

$$\hat{\sigma}^2 = \frac{1}{dn} \sum_c \sum_{\{i : y_i = c\}} \|x_i - \hat{\mu}_c\|^2$$ ⇐ pooled within class variance