More Anisotropic Gaussians

$$f(x) = \underbrace{\frac{1}{\sqrt{(2\pi)^d |\Sigma|}}\, e^{\underbrace{-(x-\mu)^T \Sigma^{-1} (x-\mu)}_{q(x)}}}_{r(q(x))}$$

note: $|\Sigma| = \det(\Sigma)$

$q(x)$ $\mathbb{R}^d \to \mathbb{R}$, quadratic

$$\Sigma = V \Lambda V^T \qquad \Lambda_{ii} = \sigma_i^2$$

eigenvalues of Σ are variances along eigenvectors

$$\Sigma^{1/2} = V \Lambda^{1/2} V^T$$ maps spheres to ellipsoids

standard deviations (widths of ellipsoids)

$$q(x) = (x-\mu)^T \Sigma^{-1} (x-\mu)$$

$$\Sigma^{-1} = V \Lambda^{-1} V^T$$ (precision matrix)

Maximum Likelihood estimation for Anisotrophic Gaussians

pts. $X_1 \ldots X_n$ classes $y_1 \ldots y_n$.

$$\hat{\Sigma}_c = \frac{1}{n_c} \sum_{i: y_i = c} (x_i - \hat{\mu}_c)(x_i - \hat{\mu}_c)^T \Leftarrow$$ conditional covariance for pts. in class c

$\hat{\pi}_c$, $\hat{\mu}_c$: Same as before | $\hat{\Sigma}_c \succeq 0$ (true normal is PD — no 0 variance dims)

$$\hat{\Sigma} = \frac{1}{n} \sum_{c} \sum_{i: y_i = c} (x_i - \hat{\mu}_c)(x_i - \hat{\mu}_c)^T \Leftarrow$$ pooled within class covariance matrix

QDA

→ choose C that maximizes $f(X=x \mid C=c)\,\pi_c$ is equivalent to maximizing the quadratic discriminant fn.

$$Q_c(x) = \ln\left[\left(\sqrt{2\pi}\right)^2 f_c(x)\,\pi_c\right]$$

Multivariate gaussian pdf

$$= -\frac{1}{2}(x-\mu_c)^T \Sigma_c^{-1}(x-\mu_c) - \frac{1}{2}\ln|\Sigma_c| + \ln\pi_c$$

2 classes: Decision fn $Q_c(x) - Q_D(x)$ is quadratic, but may be indefinite $\Rightarrow$ Bayes boundary quadratic

$$P(Y=c \mid X=x) = s\left(Q_c(x) - Q_D(x)\right) \quad : s(\cdot) \text{ is logistic fn}$$

LDA

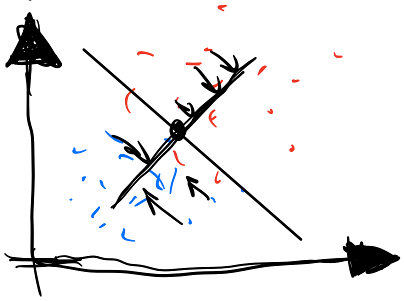
One $\hat{\Sigma}$ for all classes

$$Q_c(x) - Q_D(x) = \underbrace{(\mu_c - \mu_D)^T \Sigma^{-1} x}_{w^T x} \underbrace{- \frac{\mu_c^T \Sigma^{-1}\mu_c - \mu_D^T \Sigma^{-1}}{2} + \ln\pi_c - \ln\pi_D}_{+\ \alpha}$$

Choose class C that maximizes $\mu_c^T \Sigma^{-1}\mu_c - \frac{1}{2}\mu_c^T\Sigma^{-1}\mu_c + \ln\pi_c$

2 classes: $w^T x + \alpha = 0$ ⋮ $P(Y=c \mid X=x) = S(w^T x + \alpha)$

- LDA → project pts onto line and cut in half



for 2 classes:

- LDA has $d+1$ parameters (w, α)
- QDA has $\frac{d(d+3)}{2} + 1$ parameters

- QDA likely to overfit, LDA to underfit
- validation
- with added features, LDA can give nonlinear boundaries; QDA nonnegative
- can't get true decision boundary (b/c MLE)
- real data may not be perfectly Gaussian (prob not actually)

Terminology

Let X be $n \times d$, sample pts. one row (design matrix)

Each row i of X is a sample pt. X_i^T.

Centering X: $X \to \dot{X}$ ($-\mu^T$ every row)

Let R be uniform dist on sample pts. Sample cov matrix

$\mathrm{Var}(R) = \frac{1}{n} \dot{X}^T \dot{X}$ ← Centered X

decorrelating $\dot{X}$: $Z = \dot{X} V$, $\mathrm{Var}(R) = V \Lambda V^T$

$\mathrm{Var}(Z) = \Lambda$

__Sphering__ $\dot{X}$: $W = \dot{X}\,Var(R)^{-1/2}$

__Whitening__ $\dot{X}$: center + sphere $X \to W$

↳ W has cov. matrix I