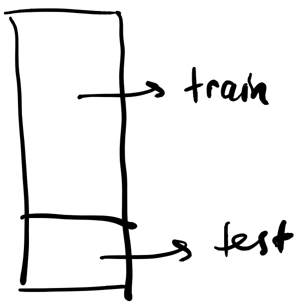
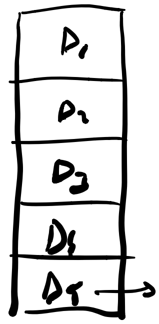
Recap of kNN, hold out, cross validation

LDA (Bayesian LDA)



hold out



D_1 test $D_2, D_3, \ldots D_5$ train

D_2 test $D_1, D_3 \ldots$ train

$\vdots$

5-fold cross validation

LDA: (Bayesian)

$(X_1, Y_1), \ldots, (X_n, Y_n)$ In LDA $X \mid Y=k \sim N(\mu_k, \Sigma)$

$$g_k(x) = \phi(\underline{x}; \underline{\mu_k}, \underline{\underline{\Sigma}})$$

$$= \frac{1}{(2\pi)^{d/2} \sqrt{|\Sigma|}} \cdot \exp\left(-\frac{1}{2}(x-\mu_k)^T \Sigma^{-1} (x-\mu_k)\right)$$

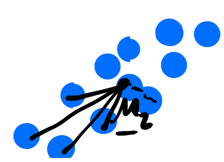
Multivariate Gaussian PDF.

LDA is the classifier obtained by plugging in the following

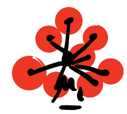
$$\hat{\pi}_k = \frac{n_k}{n} \quad \text{where} \quad n_k = \left|\{i : Y_i = k\}\right|$$

= # of class k examples

$$\hat{\mu}_k = \frac{1}{n_k} \sum_{i: y_i = k} \underline{x}_i$$



$i : y_i$ --



$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu}_{y_i})(x_i - \hat{\mu}_{y_i})^T$$

pooled covariance estimate

$\hat{\mu}$ and $\hat{\Sigma}$ are Maximum Likelihood estimates.

LDA $\quad \hat{f}(\underline{x}) = \arg\max_{k=1,2} \hat{\pi}_k \cdot \phi(\underline{x}; \hat{\mu}_k, \hat{\Sigma})$

take log

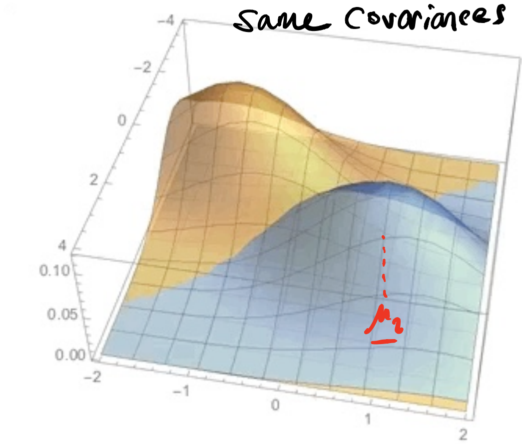
log is monotone

$f(x_1) \le f(x_2)$

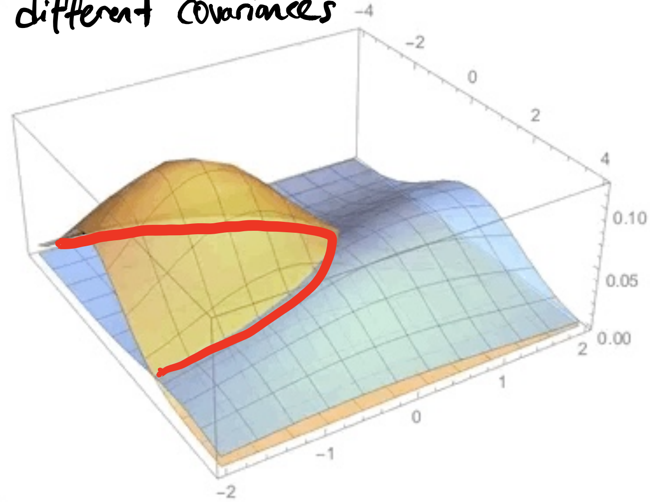
$\log f(x_1) \le \log f(x_2)$

$\arg\max f(x) = \arg\max \log f(x)$

same covariances



different covariances



$$\hat{f}(x) = \arg\max \log \hat{\pi}_k + \log \phi(x; \hat{\mu}_k, \hat{\Sigma})$$

$k = 1, 2, \dots, d$

$$\log \phi(x; \mu, \Sigma) = -\frac{d}{2} \cdot \log 2\pi - \frac{1}{2} \log |\hat{\Sigma}|$$
$$- \frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu)$$

Therefore, LDA classifier is given by

$$\log \hat{\pi}_1 - \frac{1}{2} (x - \hat{\mu}_1)^T \Sigma^{-1} (x - \hat{\mu}_1) \overset{1}{\underset{2}{\gtrless}} \log \hat{\pi}_2 - \frac{1}{2} (x - \hat{\mu}_2)^T \Sigma^{-1} (x - \hat{\mu}_2)$$

$\Updownarrow$ expand

$$\log \hat{\pi}_1 - \frac{1}{2} \cdot \left[x^T \Sigma^{-1} x - 2 x^T \Sigma^{-1} \mu_1 + \hat{\mu}_1^T \Sigma^{-1} \hat{\mu}_1 \right] \overset{1}{\underset{2}{\gtrless}}$$
$$\log \hat{\pi}_2 - \frac{1}{2} \left[x^T \Sigma^{-1} x - 2 x^T \Sigma^{-1} \mu_2 + \hat{\mu}_2^T \Sigma^{-1} \mu_1 \right]$$

$$\Downarrow$$

$$\underline{a}^T \underline{x} + b \overset{1}{\underset{2}{\gtrless}} 0 \quad \text{for some } \underline{a} \text{ and } \underline{b}.$$

Special Case : If Σ is identity (uncorrelated)

$\hat{\pi}_1 = \hat{\pi}_2 = 1/2$

$$-\frac{1}{2} \left[x^T x - 2 x^T \mu_1 + \hat{\mu}_1^T \mu_1 \right] \overset{1}{\underset{2}{\gtrless}} -\frac{1}{2} x^T x - 2 x^T \mu_2 + \hat{\mu}_2^T \mu_2$$

$$\Updownarrow$$

$$\| x - \mu_1 \|_2^2 \gtrless \| x - \mu_2 \|_2^2$$

expand $\quad (\underset{\cdot}{x} - \dot{\mu}_1)^T (\underset{\cdot}{x} - \underset{\cdot}{\mu}_1) \lessgtr (x - \mu_2)^T (x - \mu_2)$

$$x^T x - x^T \mu_1 - \mu_1^T x + \mu_1^T \mu_1 \gtrless x^T x - \mu_2^T x - x^T \mu + \mu_2^T \mu_2$$

If Σ is not identity, and we do a transformation

$x' = \Sigma^{-1/2} \cdot \underline{x}$ (matrix square root) $\Sigma = C^T C$

$\Sigma^{-1} = B^T B$