Recall **Objective**: classify $Y \in \{1, 2, \ldots, K\}$ given X.

Ingredients:
- prior prob of class k, $\pi_k = P(Y = k)$, $k = 1, \ldots, K$
- Distr. of X given Y: $f_k(x) = P(X = x \mid Y = k)$

LDA: Assumes $[X \mid Y = k] \sim N(\mu_k, \sigma^2)$

(μ_k: changes with k; σ^2: common across k)

$\Rightarrow$ discriminant functions $\delta_k(x) = \log \pi_k + x \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2}$ for $k = 1, \ldots, K$

For a new feature x_0 we predict $\hat{Y}$ is the class for which $\delta_1(x_0), \ldots, \delta_K(x_0)$ is maximized.

LDA for $p > 1$

If $\underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$ is p features. Then the LDA assumption

$$\{\underline{x} \mid Y = k\} \sim N_p(\mu_k, \Sigma)$$

→ Σ does not depend on k

The discriminant functions are

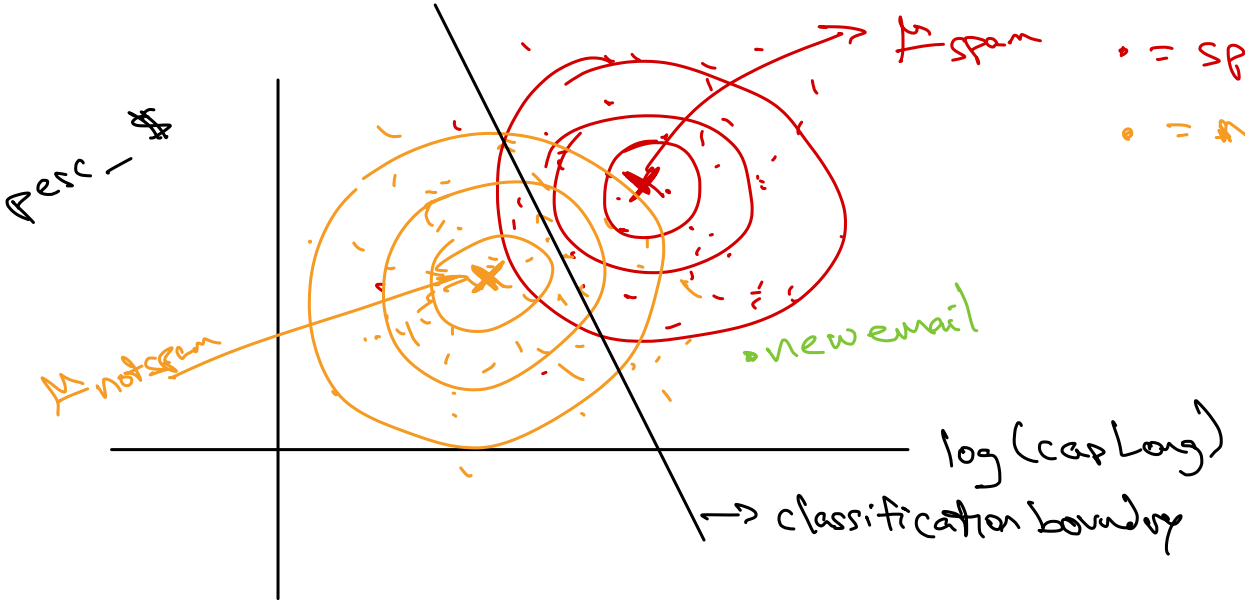
$$\delta_k(\underline{x}) = \underline{x}^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log \pi_k$$

for $k = 1, \ldots, K$

↳ still linear in $\underline{x}$

For new feature $\underline{x}_0$, $\hat{Y}$ is the class for which

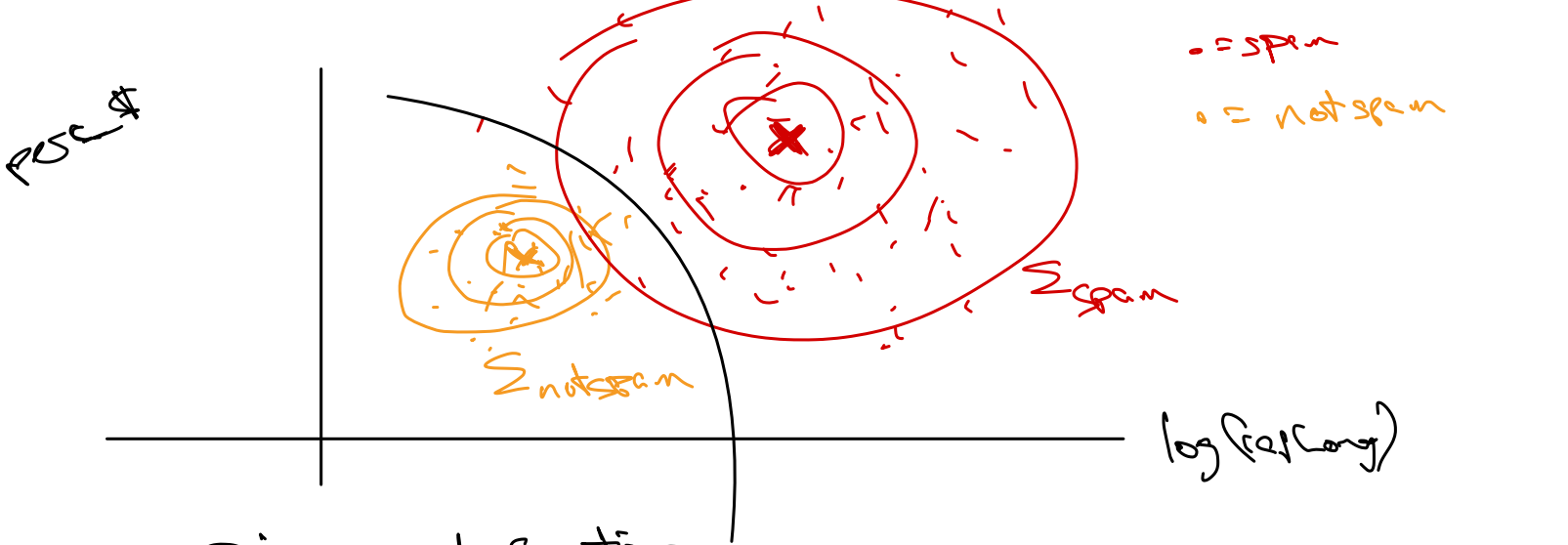
$\delta_1(\underline{x}_0), \ldots, \delta_K(\underline{x}_0)$ is biggest



Quadratic discriminant analysis (QDA)

QDA relaxes the assumption that the conditionals share a common covariance matrix, i.e.

$$[\underline{x} \mid Y = k] \sim N_p(\mu_k, \Sigma_k)$$



Discriminant functions

$$\delta_k(x) = -\frac{1}{2} x^T \Sigma_k^{-1} x + x^T \Sigma_k^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma_k^{-1} \mu_k - \frac{1}{2} \log \det \Sigma_k + \log \pi_k$$

Quadratic in x

for $k = 1, \ldots, K$

As before we compare $\delta_1(x), \ldots, \delta_K(x)$ and classify into biggest value

Remark When to select LDA vs. QDA?

- LDA when n small or is not much larger than p
- QDA when $n \gg p$ or p small

B/c Σ involves $\frac{p(p+1)}{2}$

Note LDA / QDA not appropriate for categorical features