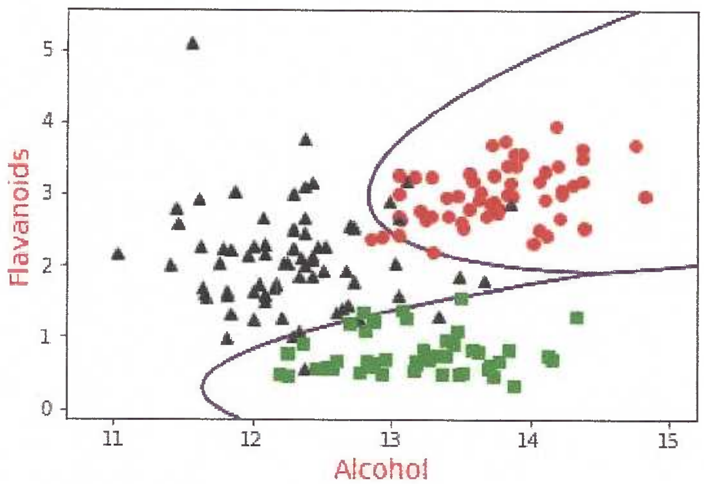
Gaussian generative models

Topics we'll cover

1. Classification using multivariate Gaussian generative modeling
2. The form of the decision boundaries results from fitting

Back to the winery data

Go from 1 to 2 features: test error goes from 29% to 8%.



With all 13 features: test error rate goes to zero.

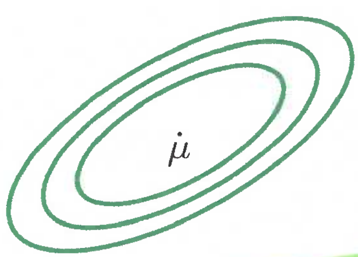
We apply Baye' rule for classificats

We fit a multivariate Gaussian to the 13 features for Winery One.
And another for winery two and another for winery 3.
Then apply Bayes rule for classificat

$\log p(x) = \text{const} + -\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)$

we see it is quadratic so the decision boundary will be quadratic

The multivariate Gaussian



$N(\mu, \Sigma)$: Gaussian in $\mathbb{R}^d$

- mean: $\mu \in \mathbb{R}^d$
- covariance: $d \times d$ matrix Σ

$$\text{Density} \quad p(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

If we write $S = \Sigma^{-1}$ then S is a $d \times d$ matrix and

$$(x-\mu)^T \Sigma^{-1}(x-\mu) = \sum_{i,j} S_{ij}(x_i - \mu_i)(x_j - \mu_j),$$

a **quadratic function** of x.

Binary classification with Gaussian generative model

if we take Deriv $P(x)$ then log, we set

- Estimate class probabilities π_1, π_2
- Fit a Gaussian to each class: $P_1 = N(\mu_1, \Sigma_1)$, $P_2 = N(\mu_2, \Sigma_2)$

this is the decision rule

Given a new point x, predict class 1 if

$$\pi_1 P_1(x) > \pi_2 P_2(x) \quad \Leftrightarrow \quad \boxed{x^T M x + 2w^T x \geq \theta,}$$

where:

quadratic boundary $\longleftarrow$ $M = \frac{1}{2}(\Sigma_2^{-1} - \Sigma_1^{-1})$ $\longleftarrow$ difference of inverse covariance matrices

linear boundary — $w = \Sigma_1^{-1}\mu_1 - \Sigma_2^{-1}\mu_2$

and θ is a threshold depending on the various parameters.

Linear or **quadratic** decision boundary.

$\Sigma \rightarrow$ covariance matrix

$\Sigma^{-1} \Rightarrow$ inverse covariance

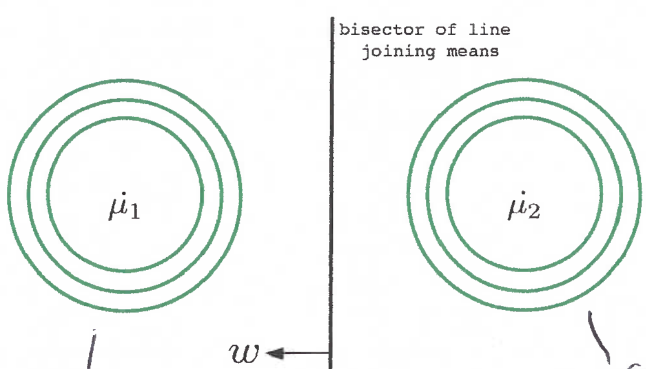
Common covariance: $\Sigma_1 = \Sigma_2 = \Sigma$

Linear decision boundary: choose class 1 if

$$x \cdot \underbrace{\Sigma^{-1}(\mu_1 - \mu_2)}_{w} \geq \theta.$$

Example 1: Spherical Gaussians with $\Sigma = I_d$ and $\pi_1 = \pi_2$.

$\pi_1 = \pi_2$ bisects

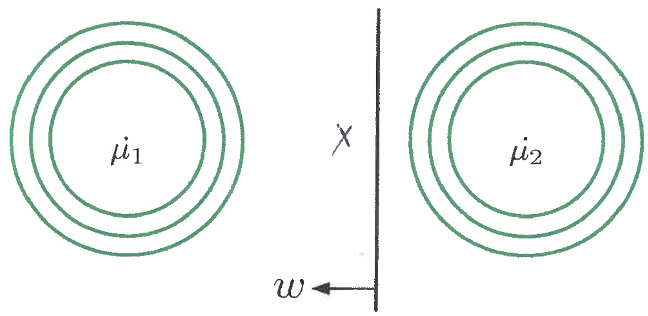


gaussian

gaussian

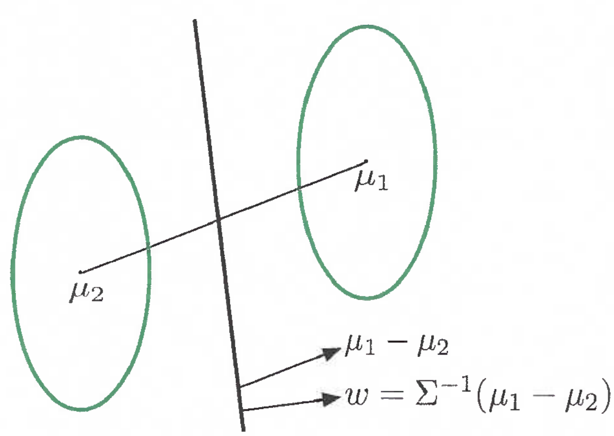
decision boundary – linear

Example 2: Again spherical, but now $\pi_1 > \pi_2$. $\Pi_1 > \Pi_2$



decision boundary gets
→ shifted slightly to the right

Example 3: Non-spherical. Same covariance matrix



slightly skewed

Classification rule: $w \cdot x \geq \theta$

- Choose w as above
- Common practice: fit θ to minimize training or validation error

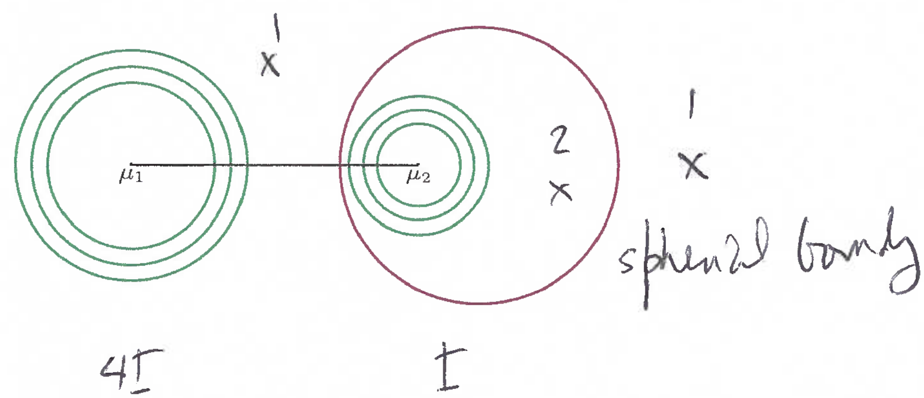
Different covariances: $\Sigma_1 \neq \Sigma_2$

Different covariance matrix so quadratic & not linear as before

Quadratic boundary: choose class 1 if $x^T M x + 2w^T x \geq \theta$, where:

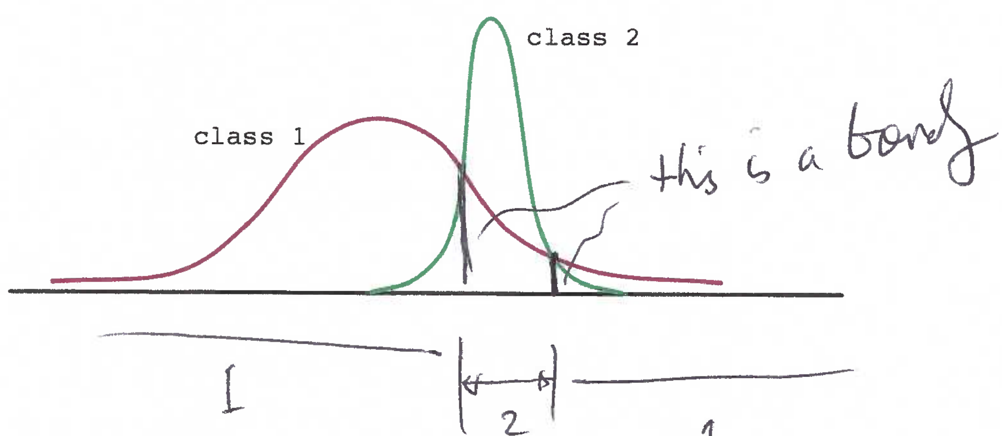
$$M = \frac{1}{2}(\Sigma_2^{-1} - \Sigma_1^{-1})$$
$$w = \Sigma_1^{-1}\mu_1 - \Sigma_2^{-1}\mu_2$$

Example 1: $\Sigma_1 = \sigma_1^2 I_d$ and $\Sigma_2 = \sigma_2^2 I_d$ with $\sigma_1 > \sigma_2$



Example 2: Same thing in 1-d. $\mathcal{X} = \mathbb{R}$.

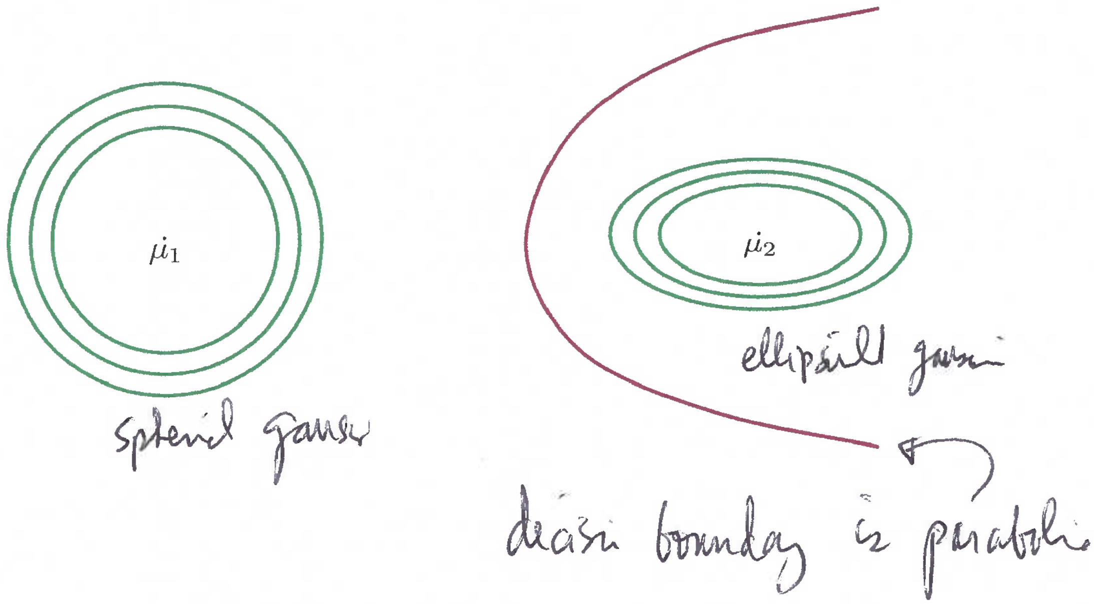
shown in 1 dimension



$\Sigma_1 \neq \Sigma_2$

the decision boundary in 1 dimension makes it hard to visualize as spherical boundary in/as shown in 2 dimension for different covariance

Example 3: A parabolic boundary.



Multiclass discriminant analysis

k classes: weights π_j, class-conditional densities $P_j = N(\mu_j, \Sigma_j)$.

Each class has an associated **quadratic** function

$$f_j(x) = \log\left(\pi_j P_j(x)\right)$$

To classify point x, pick $\arg\max_j f_j(x)$.

If $\Sigma_1 = \cdots = \Sigma_k$, the boundaries are **linear**.

all covariance matrices are equal then decision bndy is linear