Lecture 11 Generative Modeling

Recap

- Expected Value
- Correlation
- Multivariate Gaussian distributions
- Intuition behind the multivariate Gaussian formula
- Mahalanobis distance
- Minimum Covariance Determinant

Classification problem

- You have a bottle of wine with missing label
- Is it from winery 1, 2 or 3?

Data set

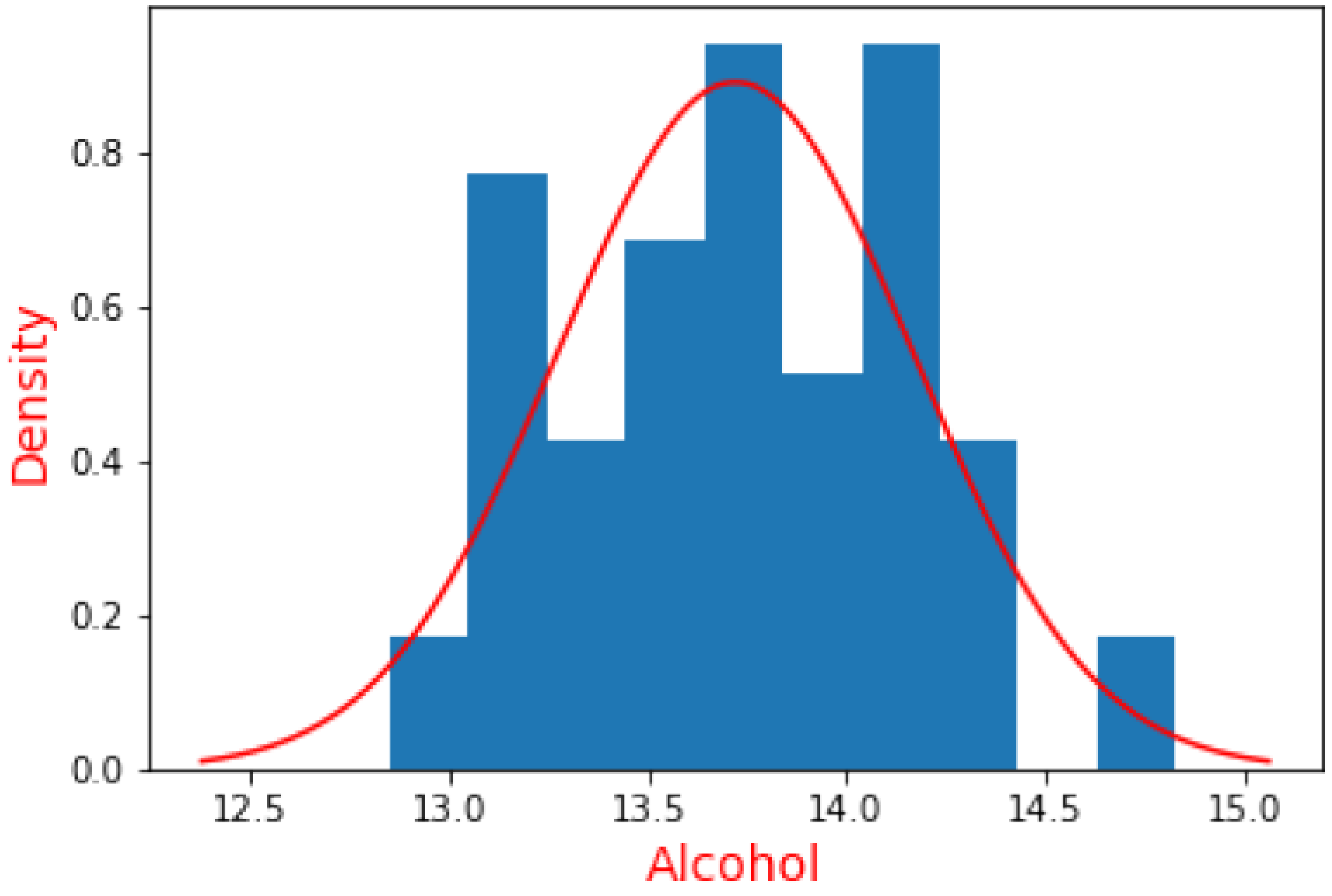
- Training set: records for 130 bottles
  - Winery 1: 10%
  - Winery 2: 50%
  - Winery 3: 40%
- 13 features
  - Alcohol, Malic acid, Flavanoids etc.
- Test set: records of 48 bottles

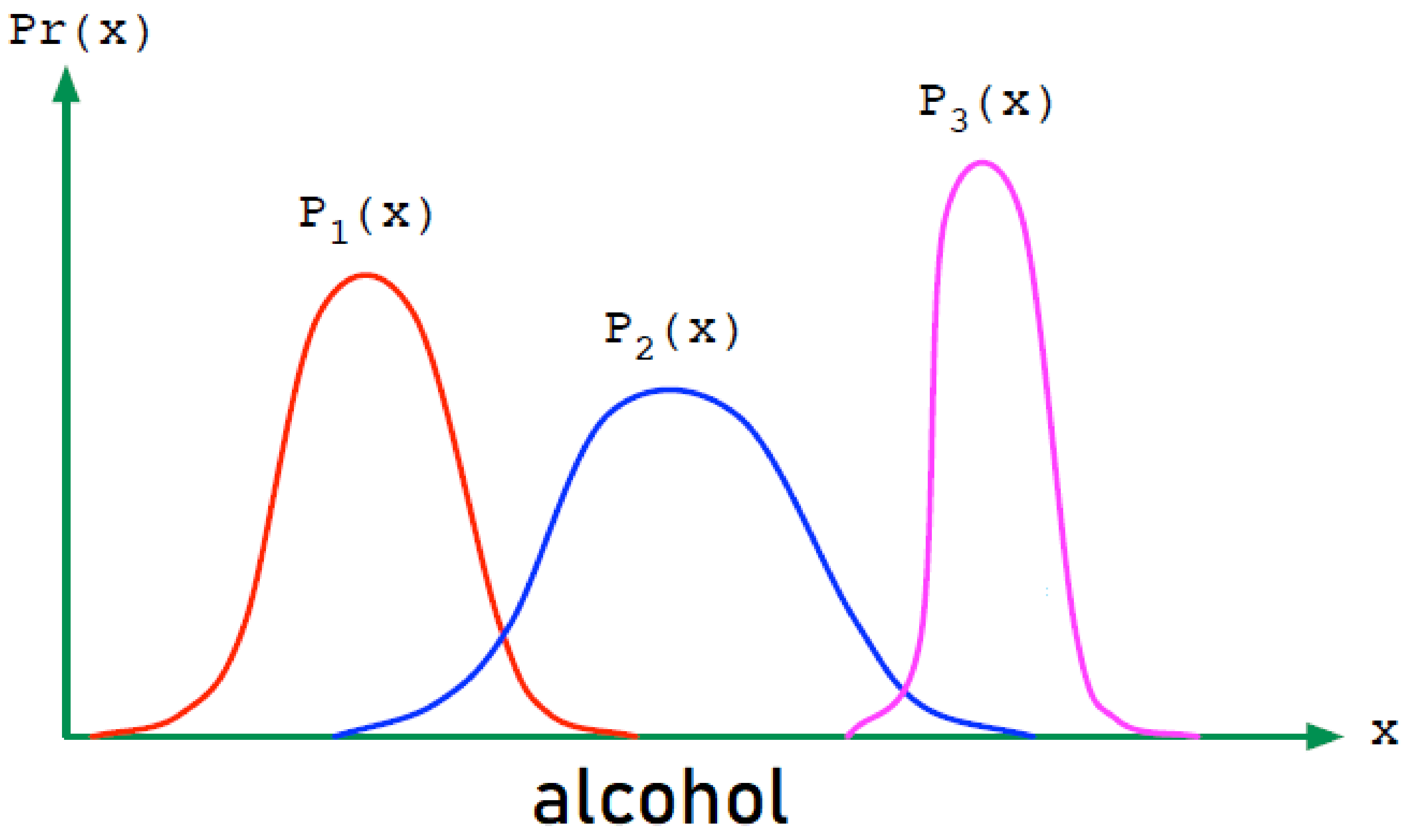
Univariate generative approach: Fit Gaussian

- Find mean and variance for each winery
- Fit 3 different Gaussian distributions for 3 wineries

$$X \sim \mathcal{N}(\mu, \sigma^2) \qquad \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \qquad X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2) \qquad X_3 \sim \mathcal{N}(\mu_3, \sigma_3^2)$$





Univariate generative approach: baby steps

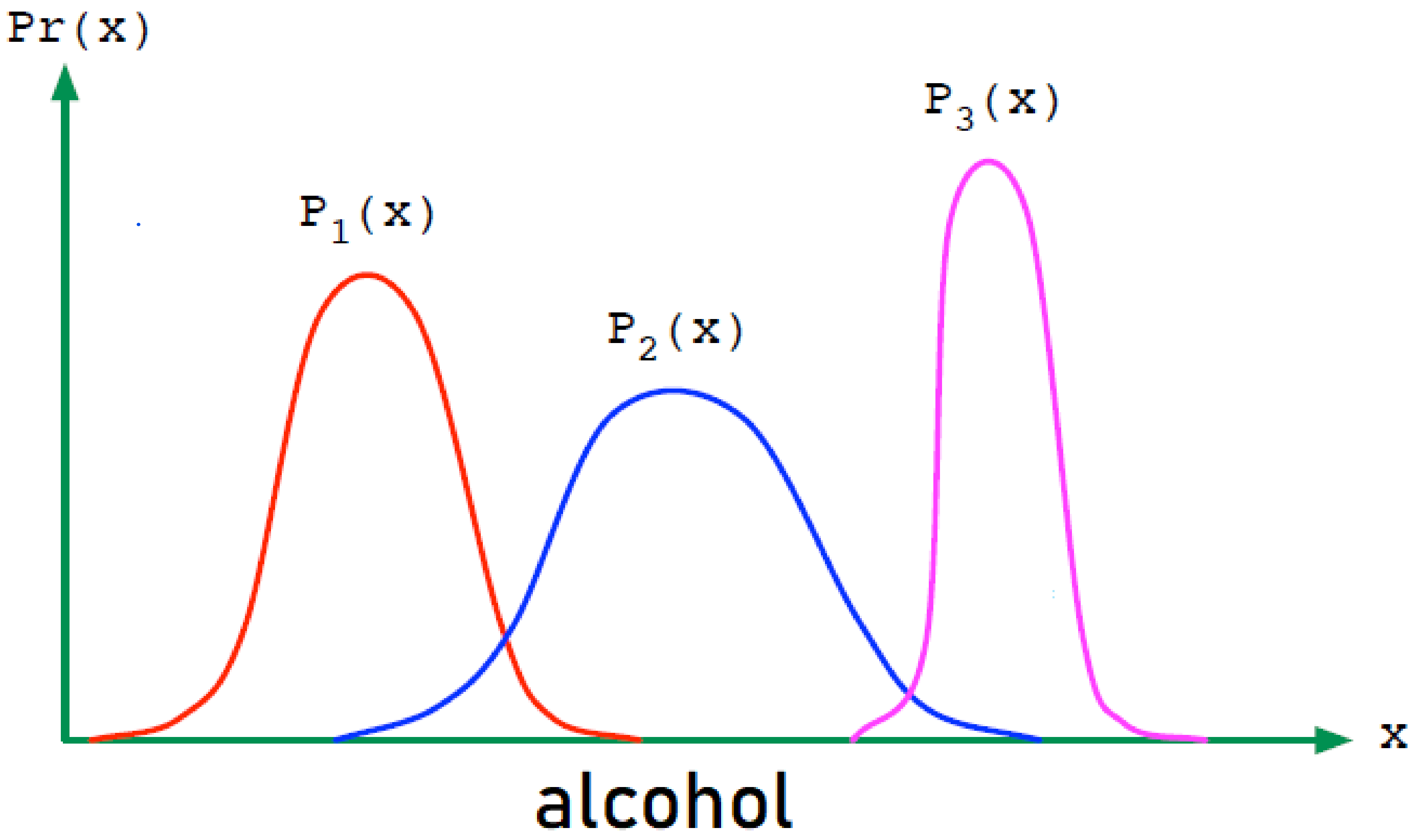
- After fitting 3 different Gaussian distributions

$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$$

$$X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$$

$$X_3 \sim \mathcal{N}(\mu_3, \sigma_3^2)$$

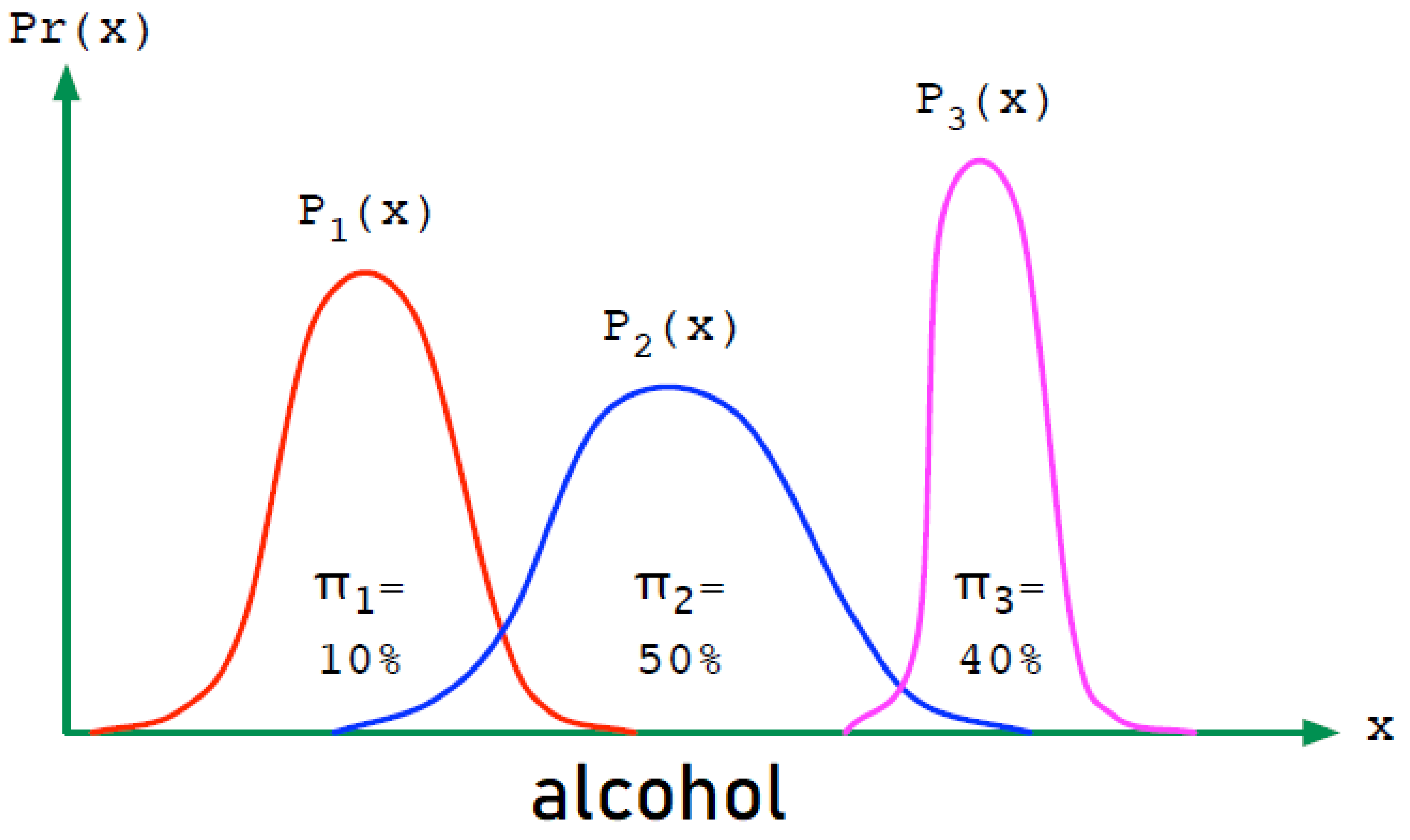
$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



- For any new x-test, plugin x-test (alcohol) into equation
- Get likelihoods for 3 wineries. Find max of 3 likelihoods

- Aside: Why do we call this likelihood instead of Probability?

Univariate generative approach: Using the prior



- Is this enough?
  - Max of Pr1(x), Pr2(x) & Pr3(x)
- What about relative percentages?
- Consider the intersection of 2 distributions

- Solution

Prior

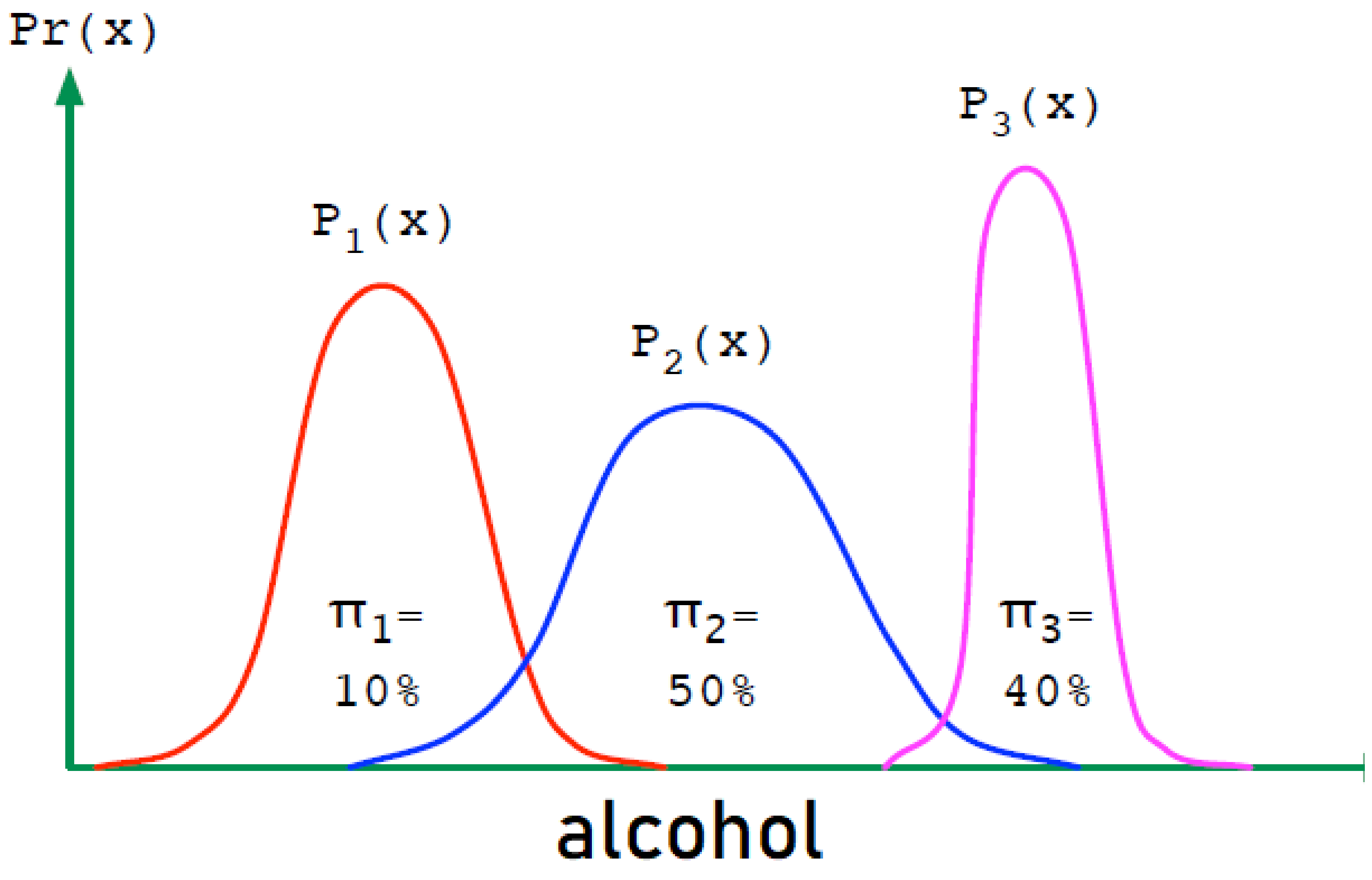
Likelihood

Max value of numerator

Posterior

$$\Pr(y=j|x) = \frac{\Pr(y=j)\Pr(x|y=j)}{\Pr(x)} = \frac{\pi_j P_j(x)}{\Pr(x)}$$

Univariate generative approach: results



- Test error = 29% with one feature
- How about we use 2 features?
- Same Bayes formula

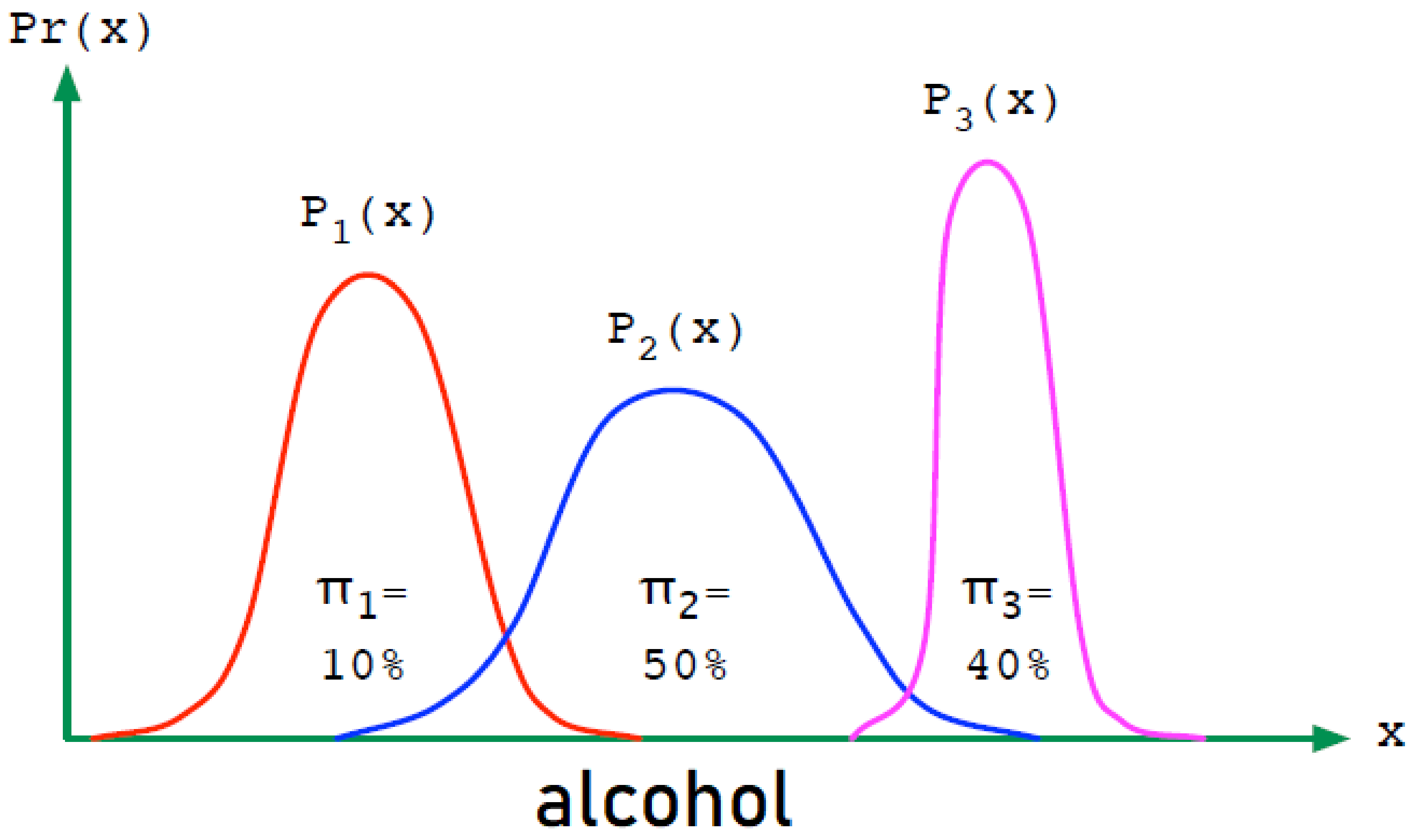
x is vector. [x1, x2, x3…]

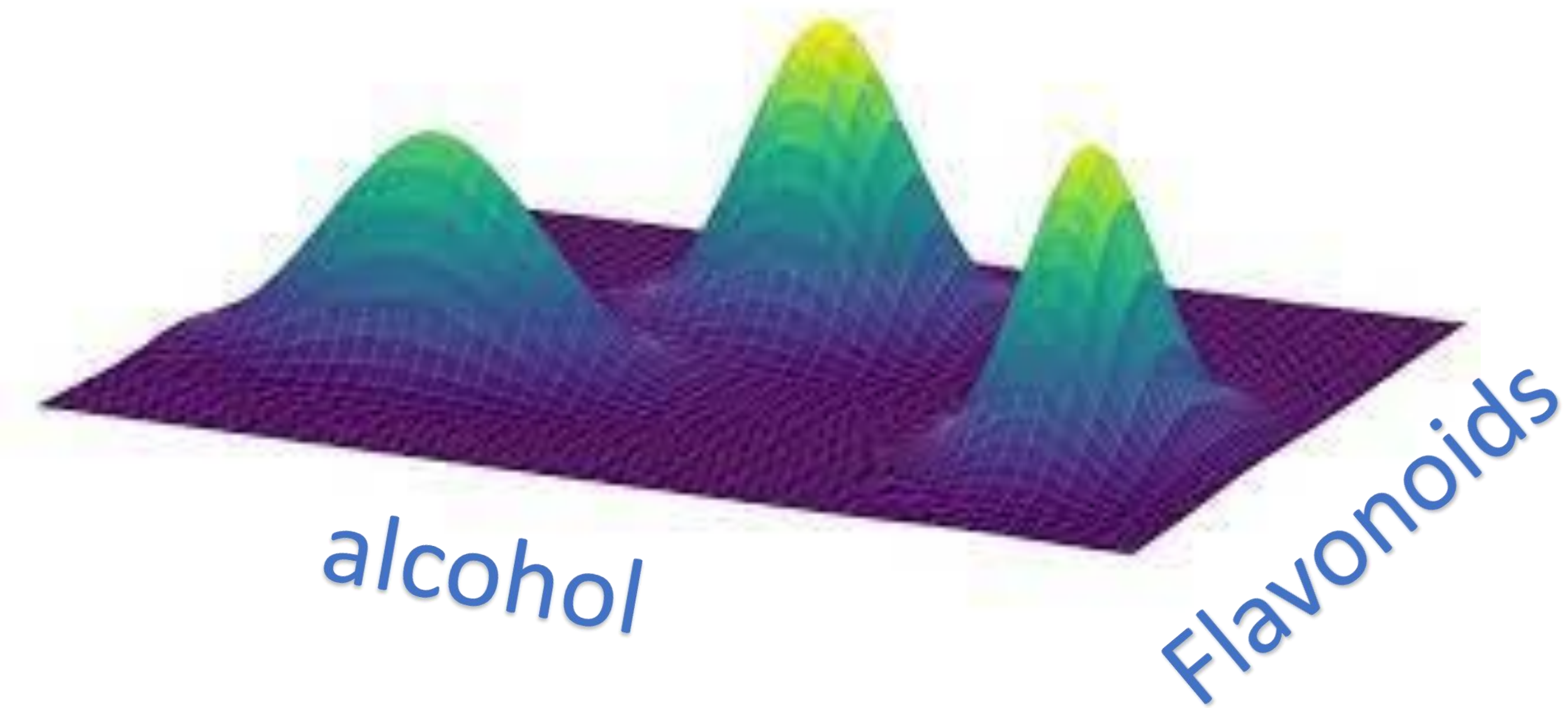
$$\Pr(y=j|x) = \frac{\Pr(y=j)\Pr(x|y=j)}{\Pr(x)} = \frac{\pi_j P_j(x)}{\Pr(x)}$$

- Fit bivariate Gaussian: x1 = Alcohol, x2 = Flavonoids

$$Pr(X) = \frac{1}{\sqrt{det(\Sigma)(2\pi)^D}} e^{\frac{-1}{2}(\mathbf{X}-\mu)^T\Sigma^{-1}(\mathbf{X}-\mu)} \qquad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

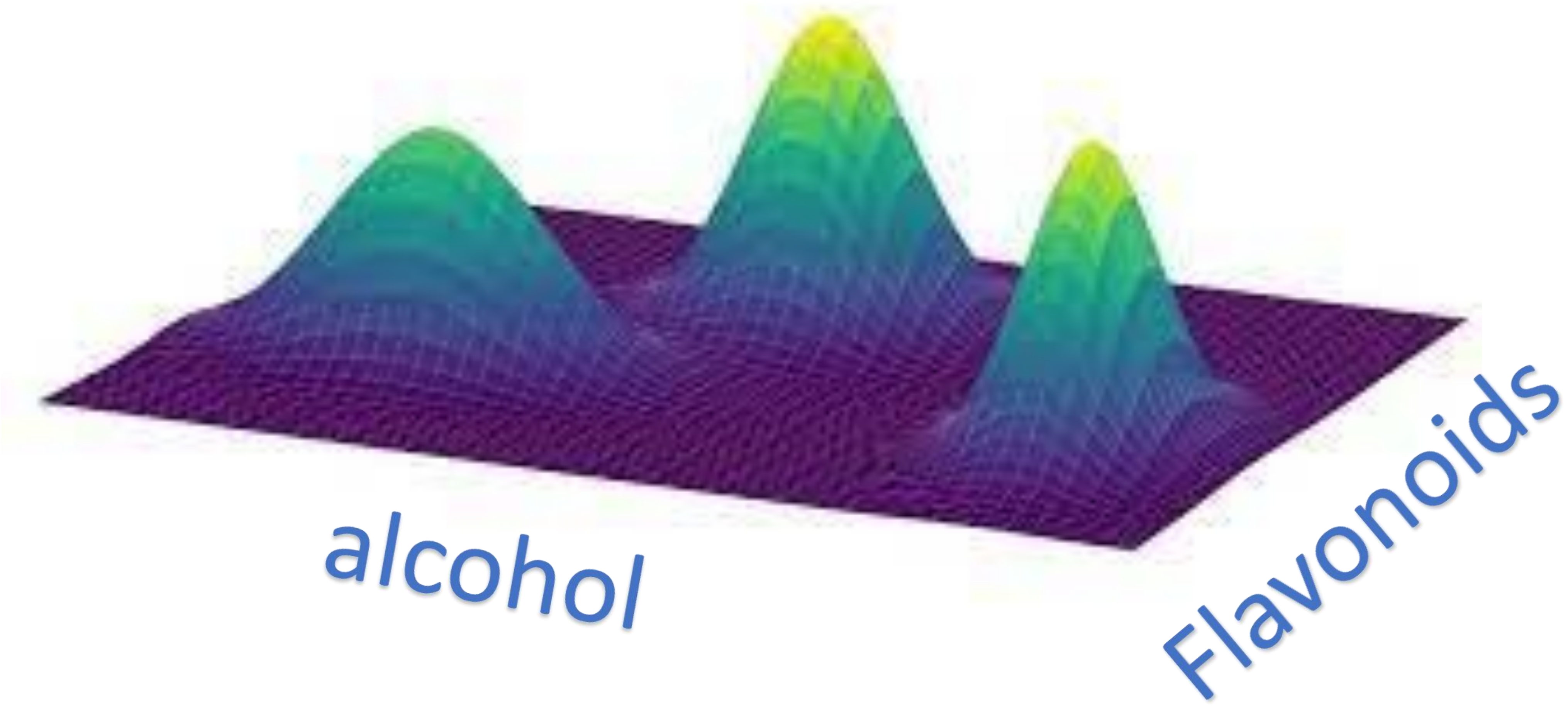
Bivariate generative approach

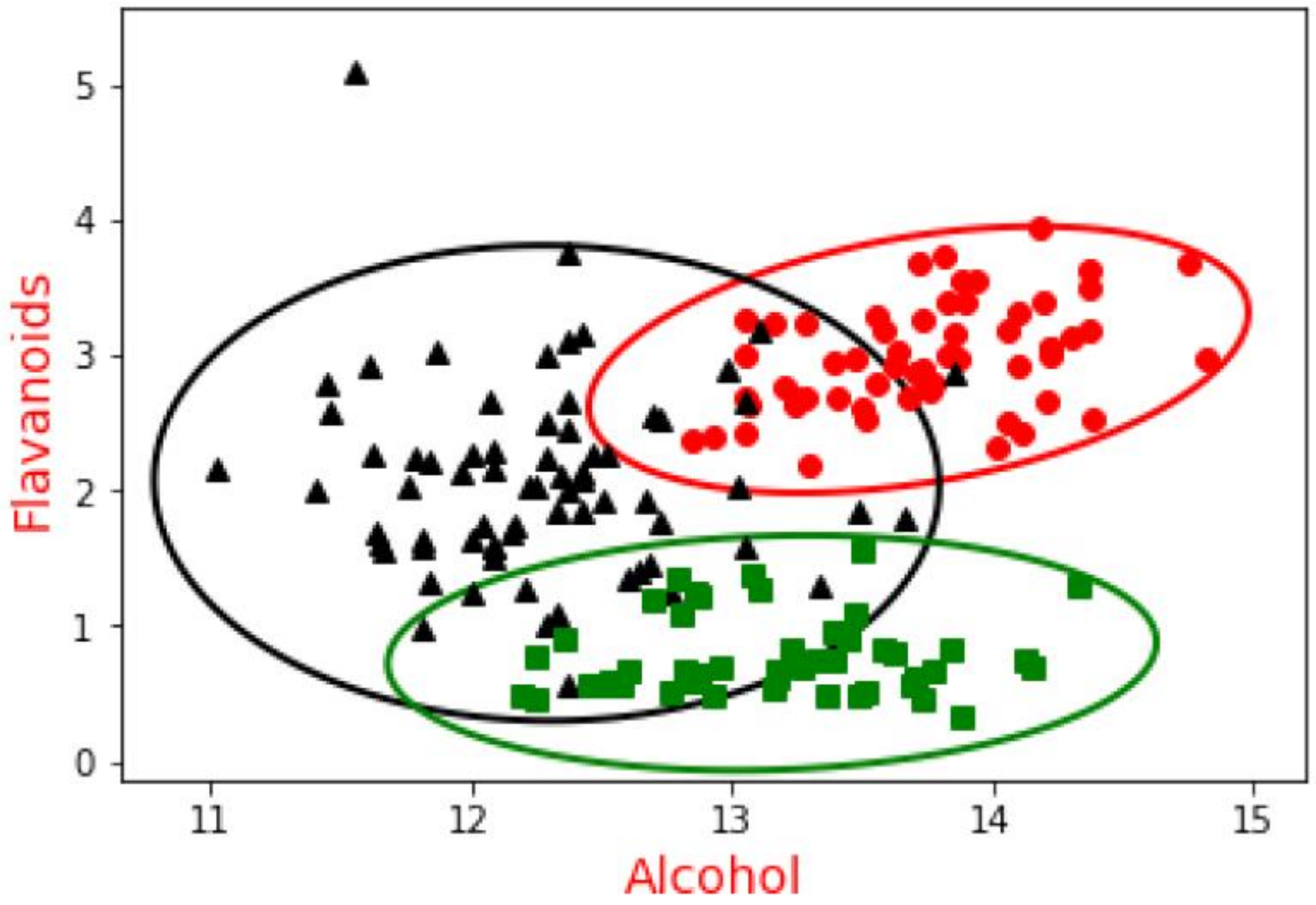




- Overlap seen in 1-D is considerably reduced in 2-D
- Better separation = Better predictions

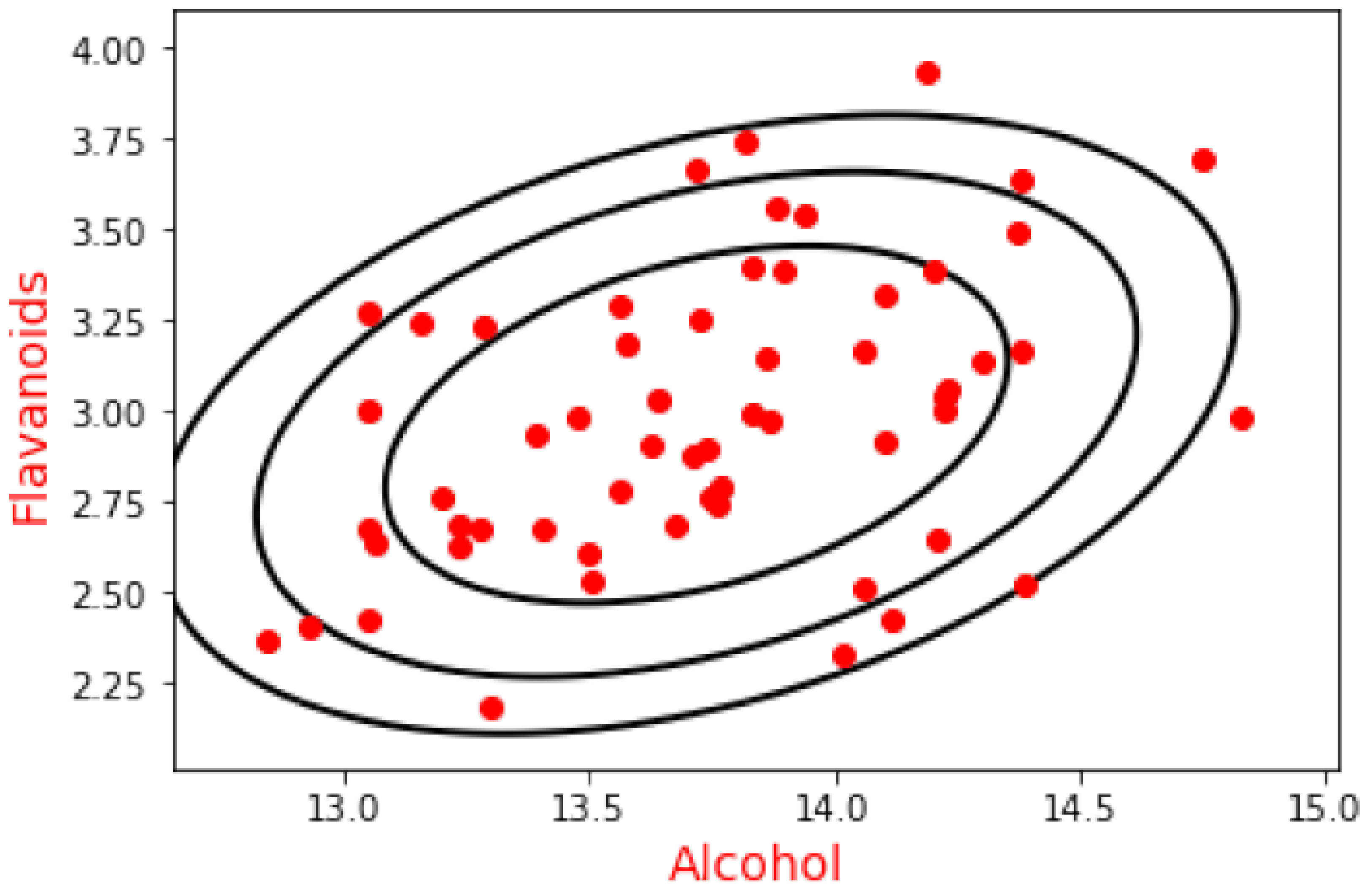
Bivariate generative approach (contd.)





- Same as Nearest Centroid model using Mahalanobis distance
- Adjusts for spread and covariance along different directions

Bivariate generative approach results



- Results: Error went from 29% to 8%
- Winery 1 class parameters

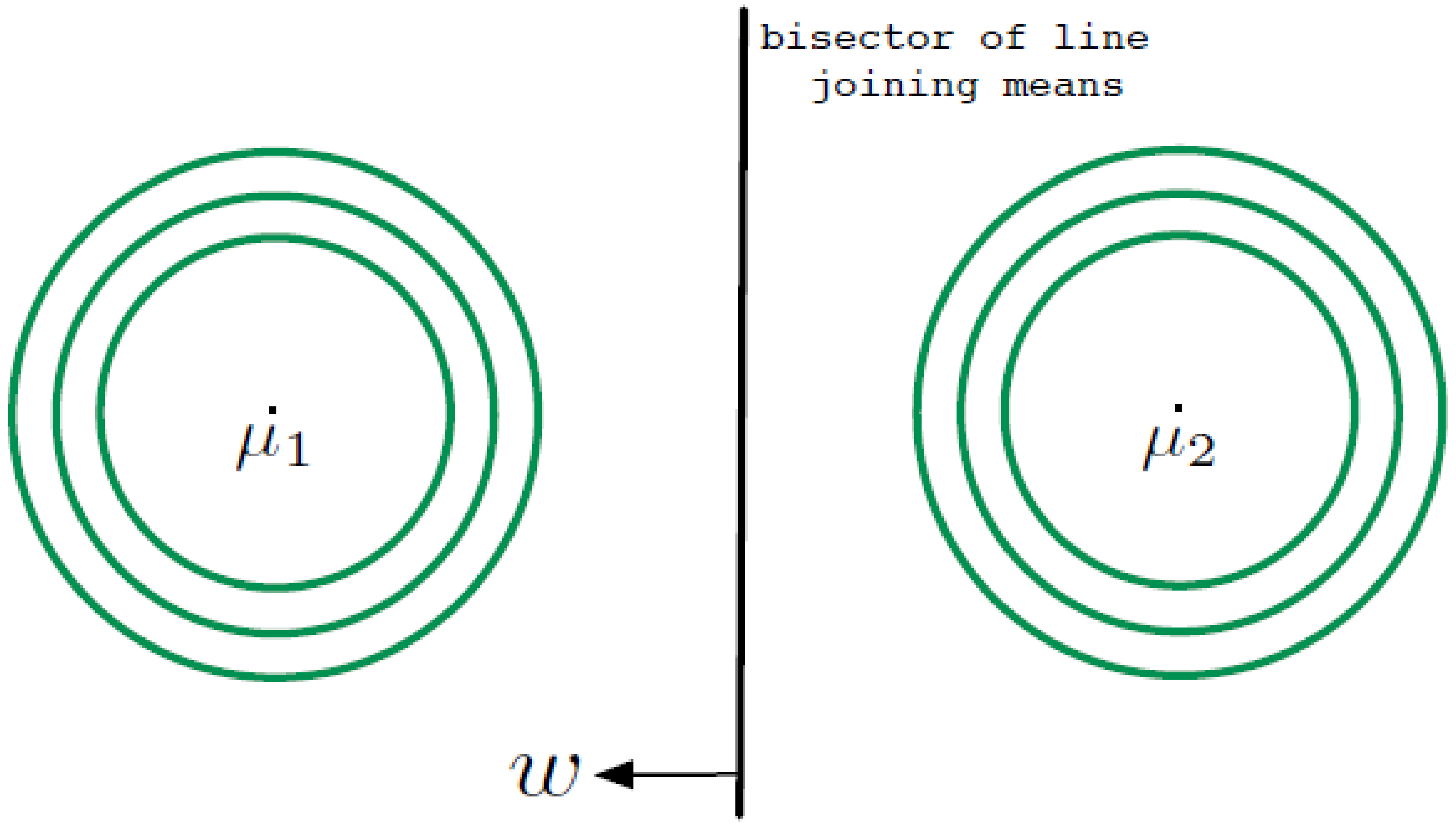
$$\mu = \begin{bmatrix} 13.7 \\ 3.0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 0.20 & 0.06 \\ 0.06 & 0.12 \end{bmatrix}$$

- Winery 2 & 3 class have different mean & covariance matrix
- How does the decision boundary look?

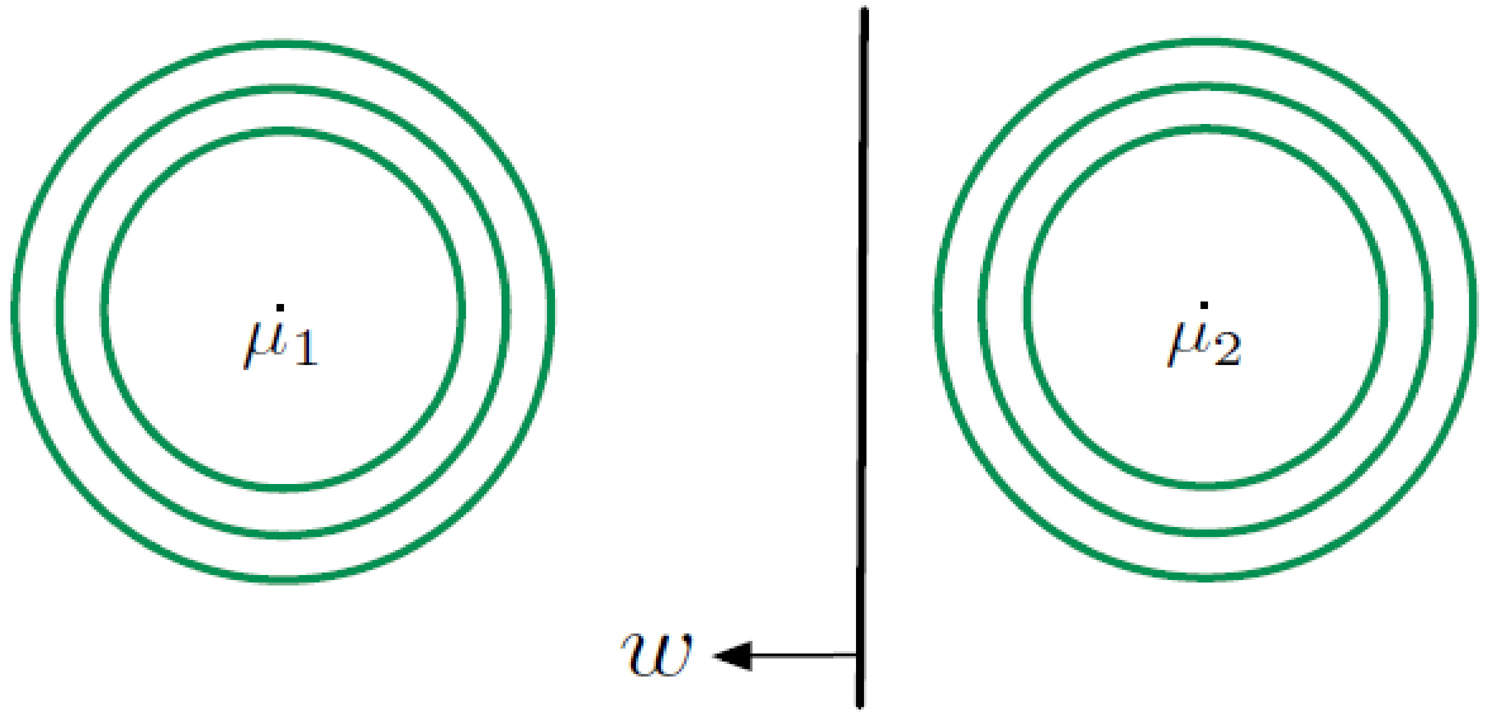
Bivariate generative: decision boundary

- Two class have same variance, same covariance
- Called: Spherical Gaussian

- Same covariance $\Sigma_1 = \Sigma_2$
- Same proportion $\pi_1 = \pi_2$

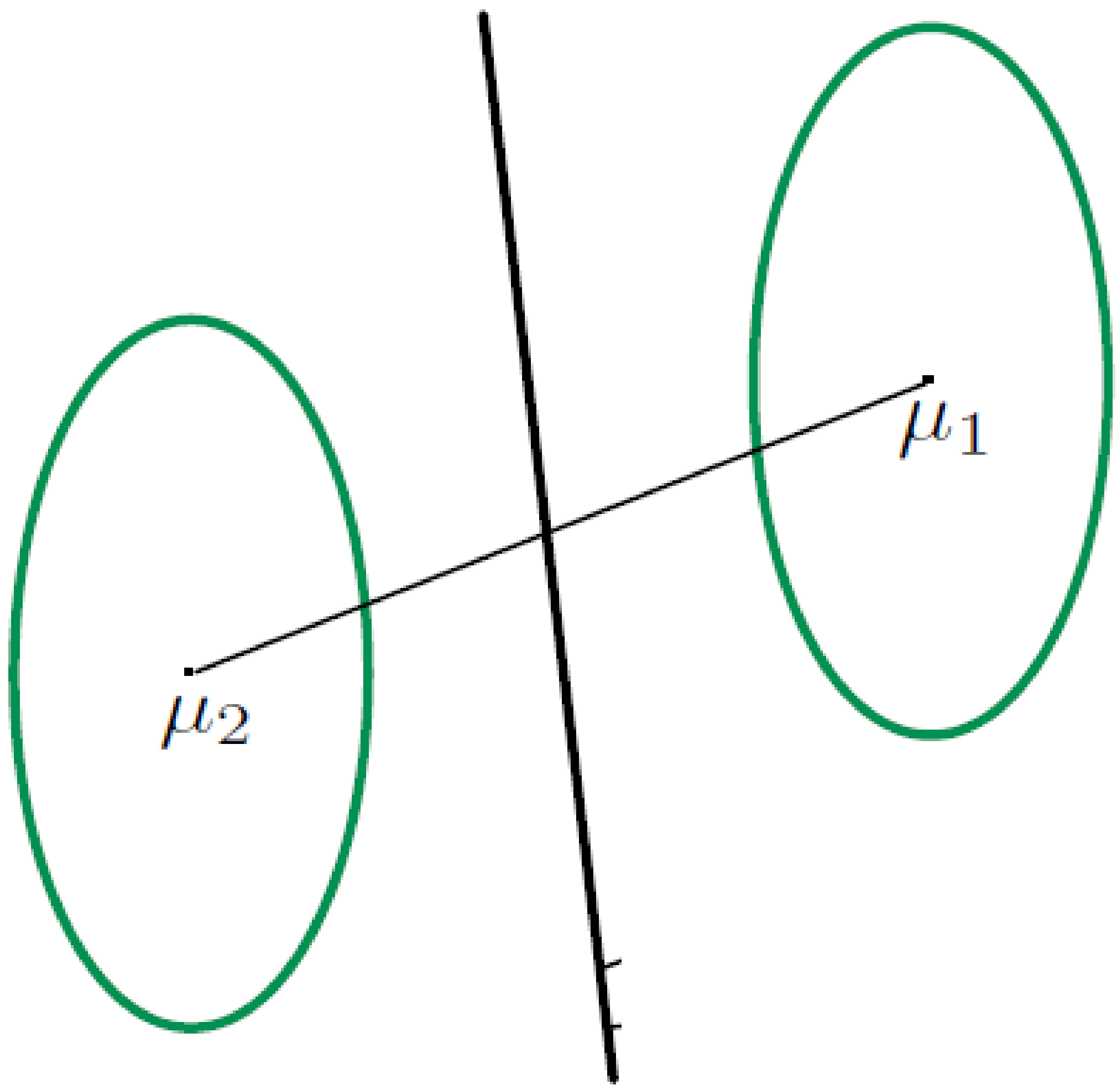


- Same covariance $\Sigma_1 = \Sigma_2$
- Diff proportion $\pi_1 > \pi_2$



Bivariate generative: decision boundary

- Two class have different variance, same covariance matrix
- Not Spherical Gaussian

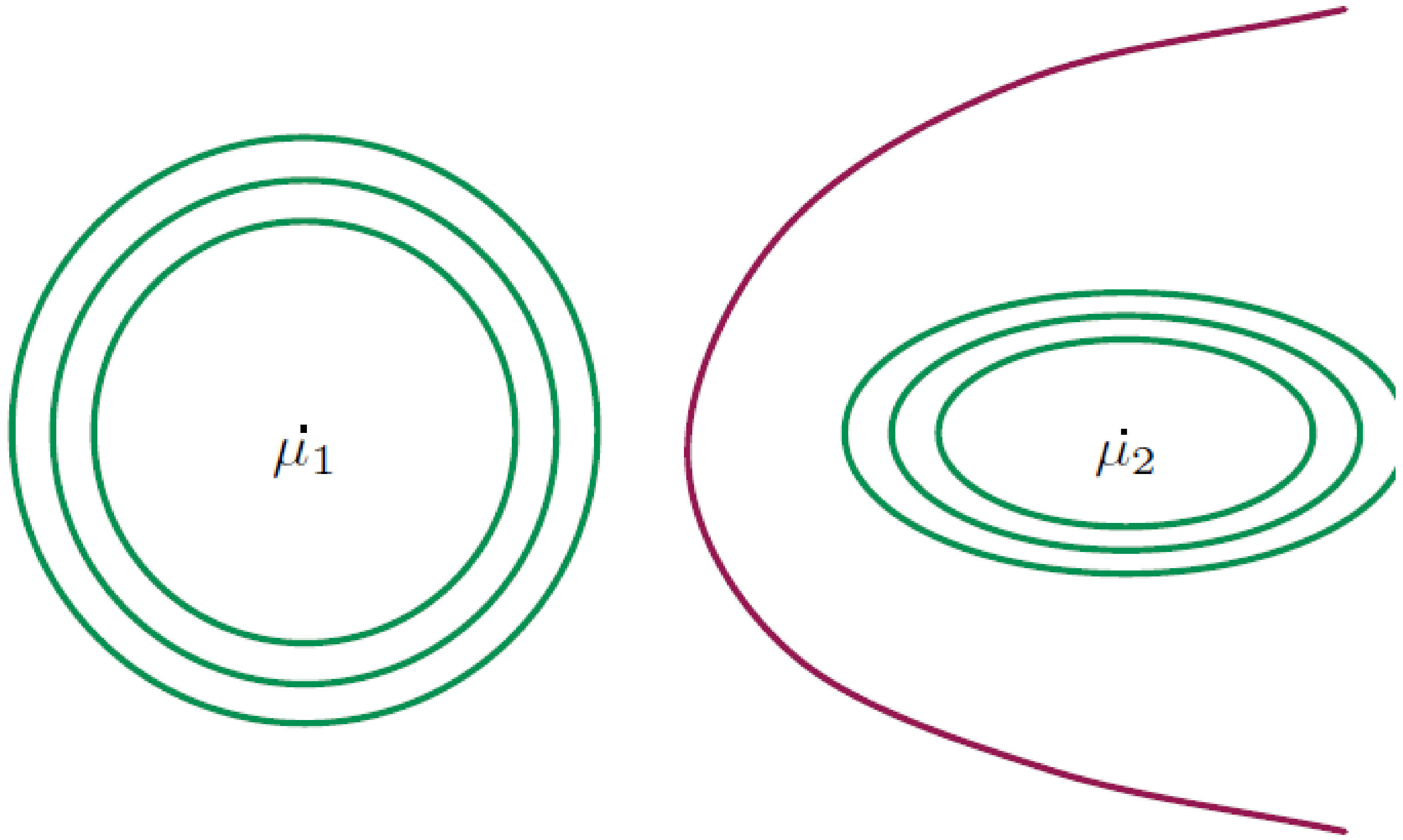


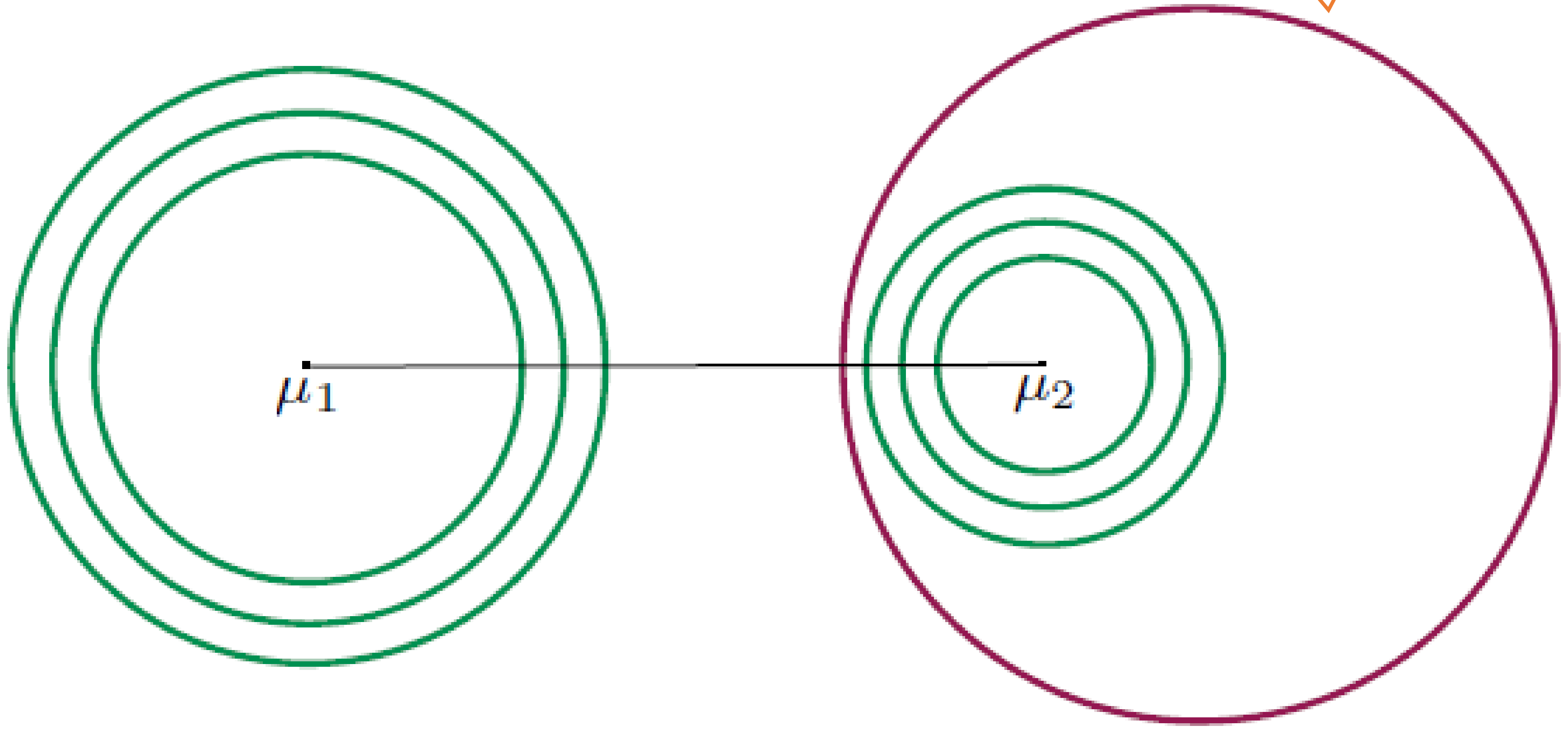
- All these linear boundaries seen till now come under Linear Discriminant Analysis (LDA)
- Second semester – MLPA subject

Bivariate generative: decision boundary

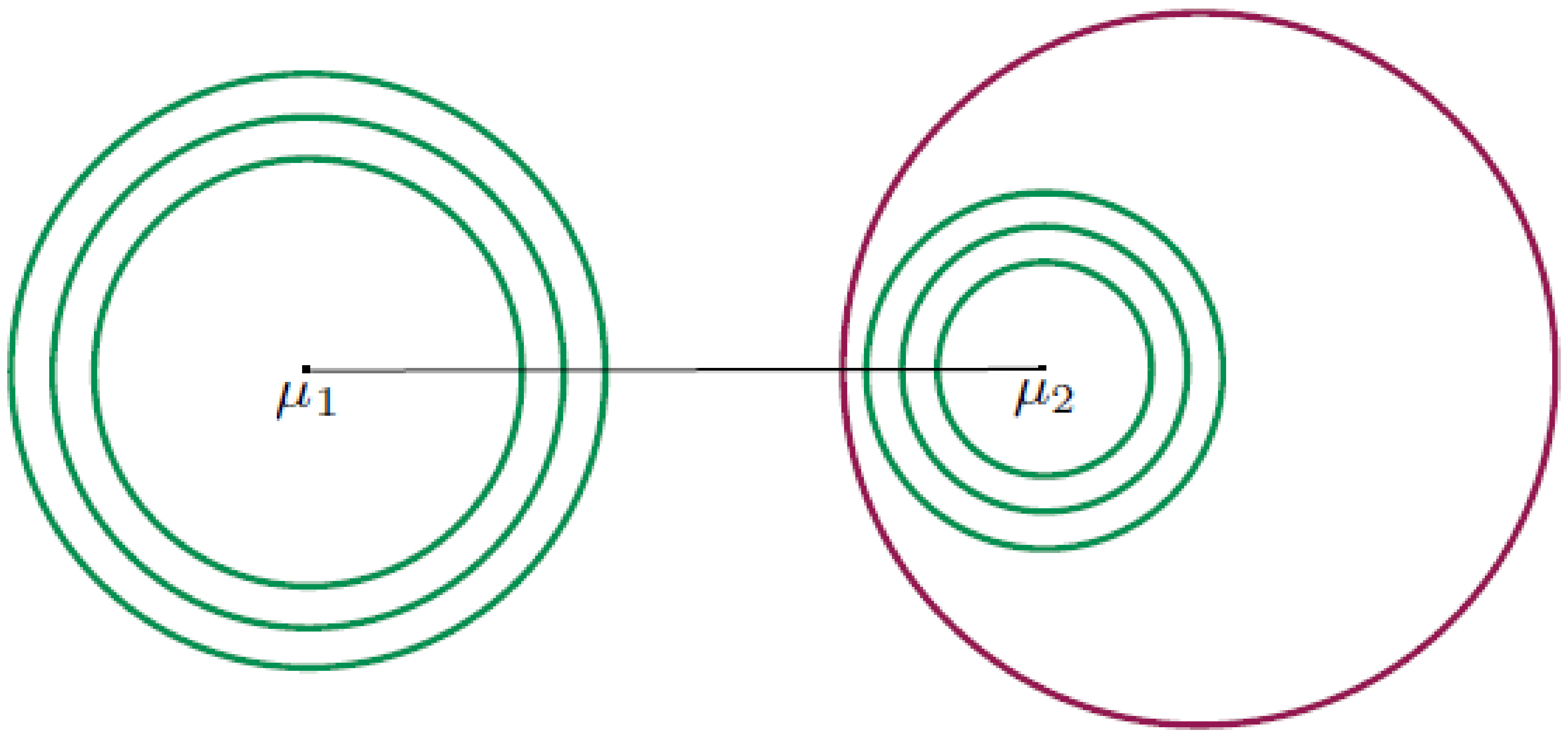
- Different covariance matrix. Quadratic decision boundary
- Quadratic Discriminant Analysis (QDA)
- Second semester – MLPA subject

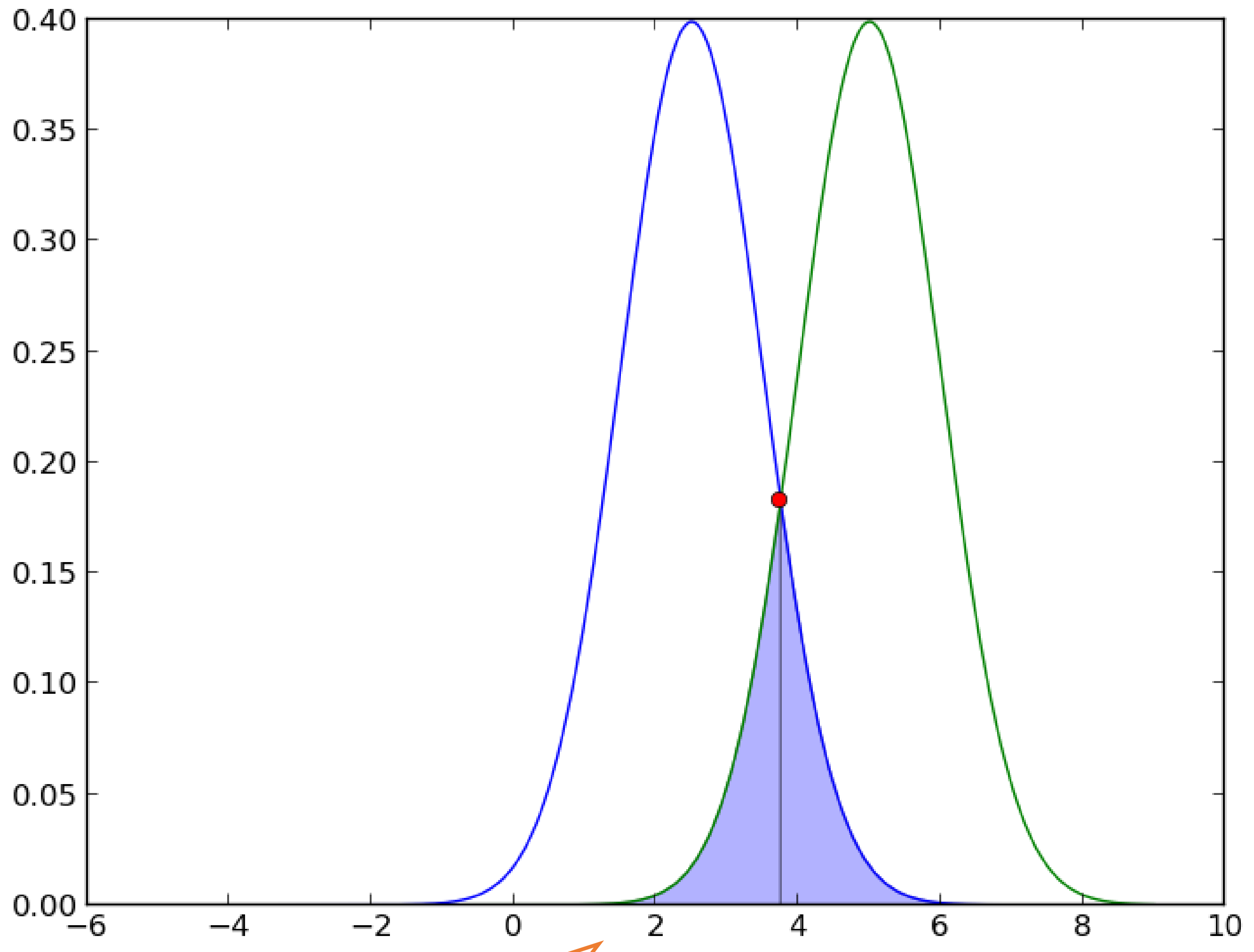
This is a bit odd. Why?



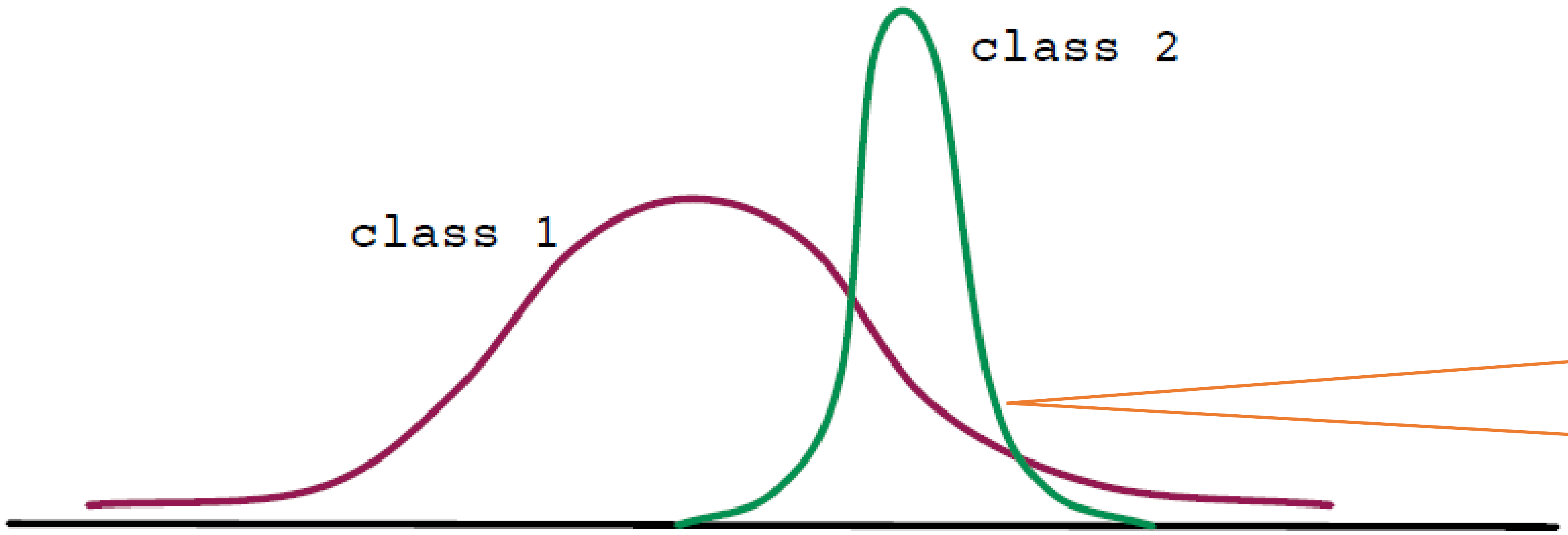


Bivariate generative: decision boundary



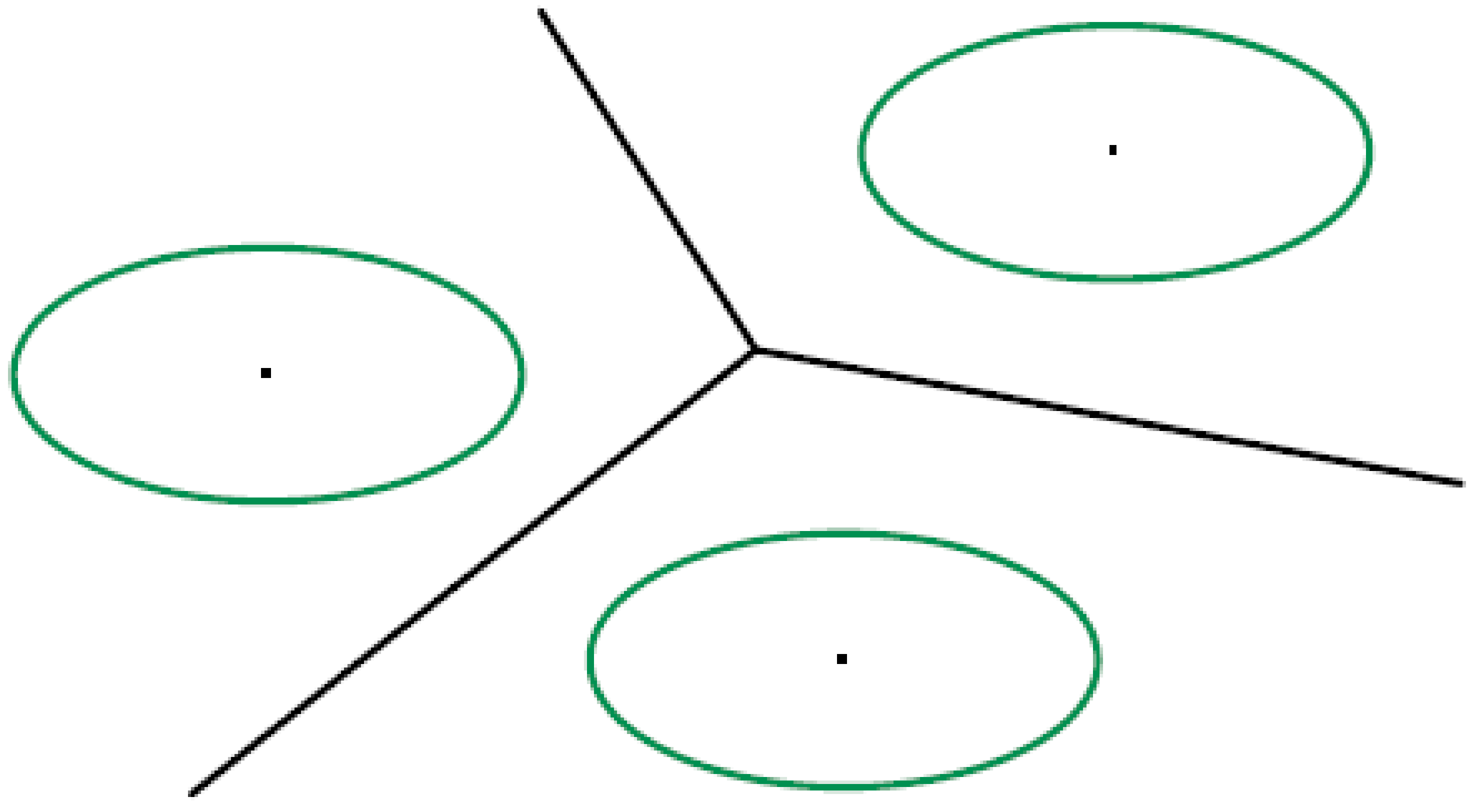


Ideal distribution for LR



If you try to fit a Logistic Regression (LR) to data like this, your performance is guaranteed to be bad

Multiclass decision boundary



Nearest Centroid v/s Gaussian comparison

Nearest Centroid	Gaussian
Eager	Eager
Batch	Batch
Parametric	Parametric
Discriminative	Generative

Naïve Bayes

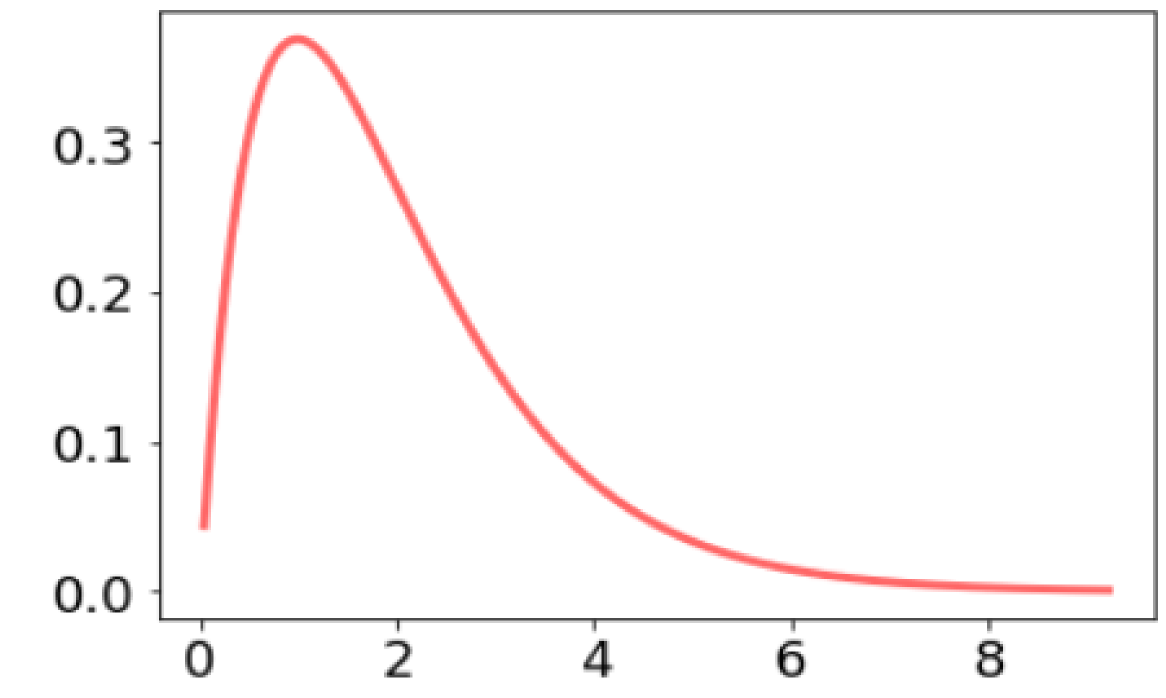
$$Pr(y|x) = \frac{Pr(x|y)Pr(y)}{Pr(x)}$$

$$Pr(y=j|x) = \frac{Pr(x|y=j)Pr(y=j)}{Pr(x)}$$

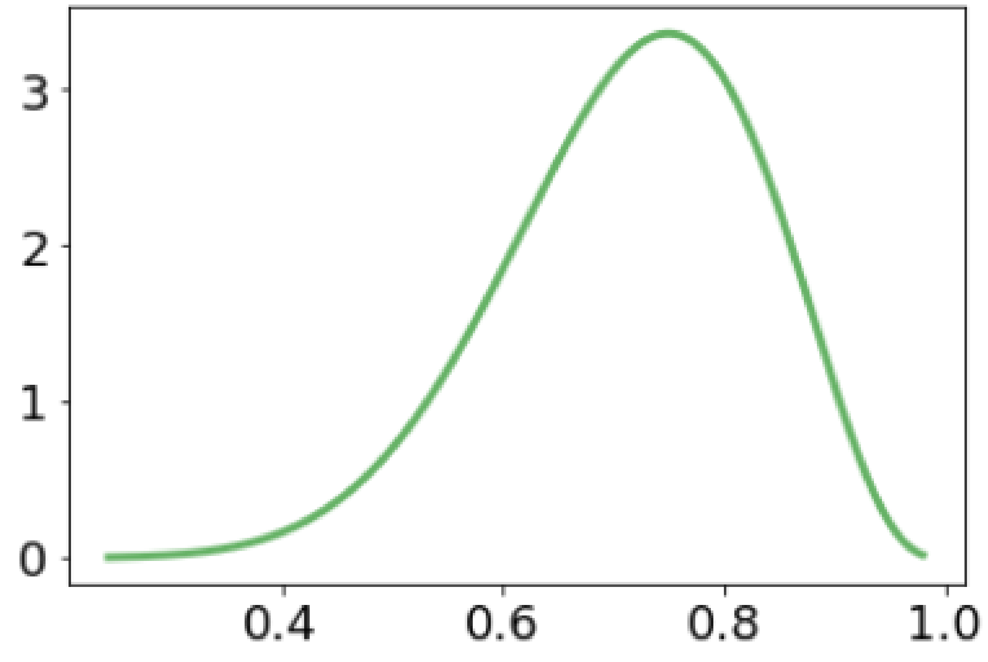
- When features are independent

$$Pr(y=j|x) = \frac{Pr(x_1|y=j)Pr(x_2|y=j)..Pr(x_n|y=j)Pr(y=j)}{Pr(x)}$$

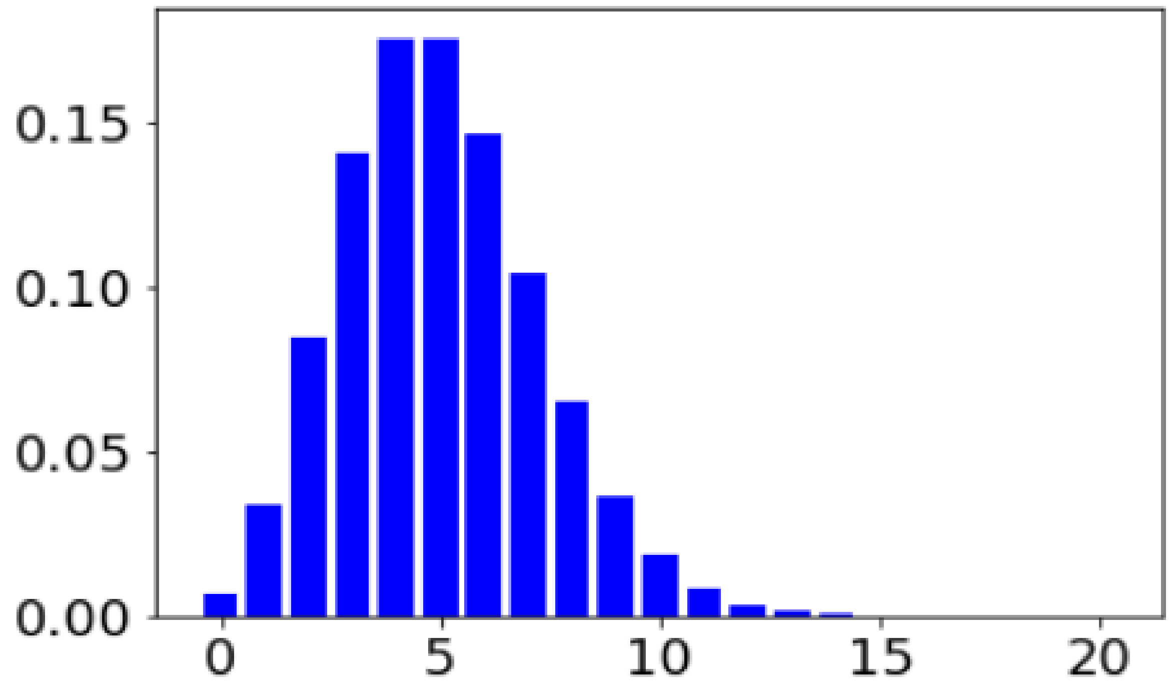
Other distributions



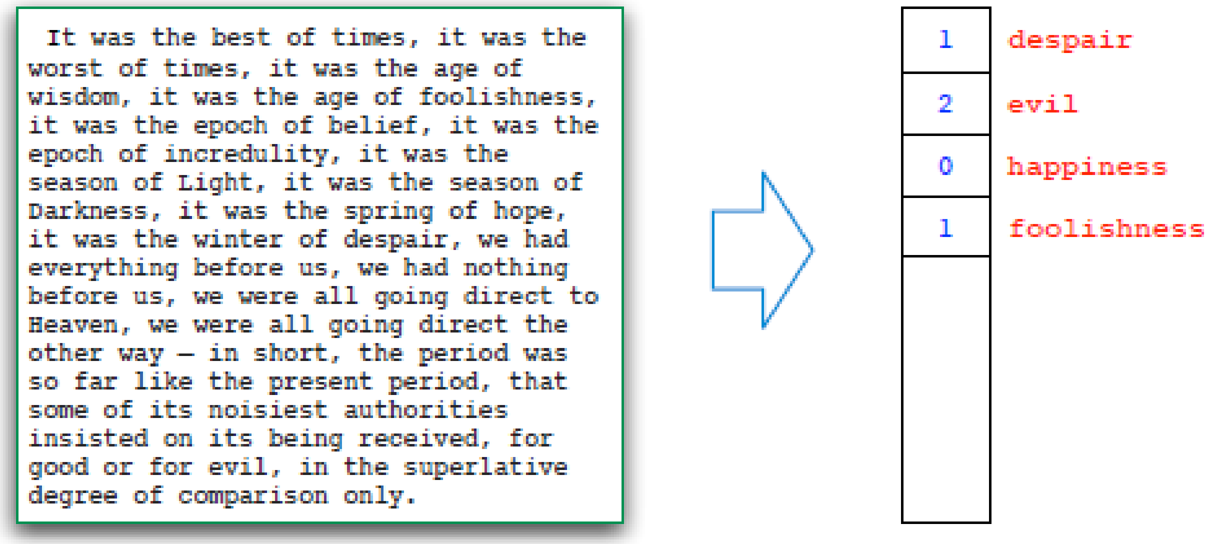
GAMMA



BETA



POISSON



CATEGORICAL


QUESTIONS

Literature Survey