Question 1

Gaussian parameters. Each of the following scenarios describes a joint distribution (x, y). In each case, give the parameters of the (unique) bivariate Gaussian that satisfies these properties.

(a) x has mean 2 and standard deviation 1, y has mean 2 and standard deviation 0.5, and the correlation between x and y is -0.5.

(b) x has mean 1 and standard deviation 1, and y is equal to 2.

Solution (a)

For a bivariate Gaussian distribution, we need to specify the following parameters:

$$\boldsymbol{\mu} = [\mu_x, \mu_y]^T$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

- Mean vector: $\boldsymbol{\mu}$

- Covariance matrix: $\boldsymbol{\Sigma}$

Let $\mu_x = 2$, $\mu_y = 2$, $\sigma_x = 1$, $\sigma_y = 0.5$, and $\rho = -0.5$

Solve for mean vector $\boldsymbol{\mu}$:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_x & \mu_y \end{bmatrix}^T = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

Solve for covariance matrix $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} 1^2 & (-0.5)(1)(0.5) \\ (-0.5)(1)(0.5) & 0.5^2 \end{bmatrix} = \begin{bmatrix} 1 & -0.25 \\ -0.25 & 0.25 \end{bmatrix}$$

$\therefore$ the bivariate Gaussian has parameters:

$$\mathcal{N}\left(\begin{bmatrix} 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 & -0.25 \\ -0.25 & 0.25 \end{bmatrix} \right)$$

Solution (b)

For a bivariate Gaussian distribution, we need to specify the following parameters:

$$\boldsymbol{\mu} = [\mu_x, \mu_y]^T$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

- Mean vector: $\boldsymbol{\mu}$

- Covariance matrix: $\boldsymbol{\Sigma}$

Let $\mu_x = 1$, $\mu_y = 2$, $\sigma_x = 1$, $\sigma_y = 0$

Solve for mean vector $\boldsymbol{\mu}$:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_x & \mu_y \end{bmatrix}^T = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Solve for covariance matrix $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} 1^2 & 0 \\ 0 & 0^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$\therefore$ the bivariate Gaussian has parameters:

$$\mathcal{N}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right)$$

Question 2

A generative approach is used for a binary classifcation problem (with classes +, -) and it turns out that the resulting classifier predicts + at all points x in the input space. Why might this be?

Solution

Step 1: Recall the Bayesian decision rule for generative classifiers.

In a generative approach, we model the joint distribution $p(x, y)$ by learning the class-conditional densities $p(x|y)$ and prior probabilities $p(y)$. The classification decision is made using Bayes' rule:

$$\hat{y} = \arg \max_{y \in \{+,-\}} p(y|x) = \arg \max_{y \in \{+,-\}} \frac{p(x|y)p(y)}{p(x)}$$

Since $p(x)$ is constant for both classes, the decision rule simplifies to:

$$\hat{y} = \arg \max_{y \in \{+,-\}} p(x|y)p(y)$$

Step 2: Analyze the decision boundary.

For binary classification, we predict the positive class (+) when:

$$p(x|+)p(+) > p(x|-)p(-)$$

Equivalently, taking the logarithm (which preserves the inequality since log is monotonically increasing):

$$\log p(x|+) + \log p(+) > \log p(x|-) + \log p(-)$$

Step 3: Identify possible reasons for always predicting the positive class.

There are several reasons why the inequality $p(x|+)p(+) > p(x|-)p(-)$ might hold for all x:

1. **Highly imbalanced prior probabilities:** If $p(+) \gg p(-)$, the classifier might always predict the positive class because the prior term dominates the decision, regardless of the likelihood term. This occurs when the training data contains many more positive examples than negative ones.

2. **Poor estimation of class-conditional densities:** If $p(x|+)$ is consistently overestimated or $p(x|-)$ is consistently underestimated across the input space, the classifier will favor the positive class.

3. **Model misspecification:** If the assumed parametric form of $p(x|y)$ (e.g., Gaussian) does not match the true distribution of the data, the estimated densities may lead to systematic classification errors.

4. **Feature insufficiency:** If the selected features do not provide discriminative information to separate the classes, the model might default to the more common class.

5. **Numerical issues:** In high-dimensional spaces, numerical underflow when computing likelihoods can lead to computational issues that affect the decision boundary.

Step 4: Mathematical illustration with Gaussian class-conditionals.

For a concrete example, consider a generative classifier with Gaussian class-conditionals:

$$p(x|+) \sim \mathcal{N}(\mu_+, \Sigma_+)$$
$$p(x|-) \sim \mathcal{N}(\mu_-, \Sigma_-)$$

The decision boundary is determined by:

$$\log p(x|+) + \log p(+) = \log p(x|-) + \log p(-)$$

Expanding the Gaussian log-likelihoods:

$$-\frac{1}{2}(x - \mu_+)^T \Sigma_+^{-1}(x - \mu_+) - \frac{1}{2}\log|\Sigma_+| - \frac{d}{2}\log(2\pi) + \log p(+) =$$
$$-\frac{1}{2}(x - \mu_-)^T \Sigma_-^{-1}(x - \mu_-) - \frac{1}{2}\log|\Sigma_-| - \frac{d}{2}\log(2\pi) + \log p(-)$$

If this equation has no solution for any x (i.e., one side is always greater), the classifier will always predict the same class. This can happen if:

- The means μ_+ and μ_- are very close, but $p(+) \gg p(-)$

- The covariance matrices Σ_+ and Σ_- are poorly estimated

- The dimensionality d is high relative to the sample size, leading to unreliable parameter estimates

Conclusion: A generative classifier that always predicts the positive class indicates an issue with either the data distribution (class imbalance), the model specification, or the parameter estimation process. In practice, this suggests the need to:

- Balance the training data or adjust class priors

- Try different parametric forms for the class-conditional densities

- Apply regularization to improve parameter estimation

- Consider dimensionality reduction if working in high-dimensional spaces

- Validate the model's assumptions about the data distribution

Question 3

Gaussian parameters. Each of the following scenarios describes a joint distribution (x, y). In each case, give the parameters of the (unique) bivariate Gaussian that satisfies these properties.
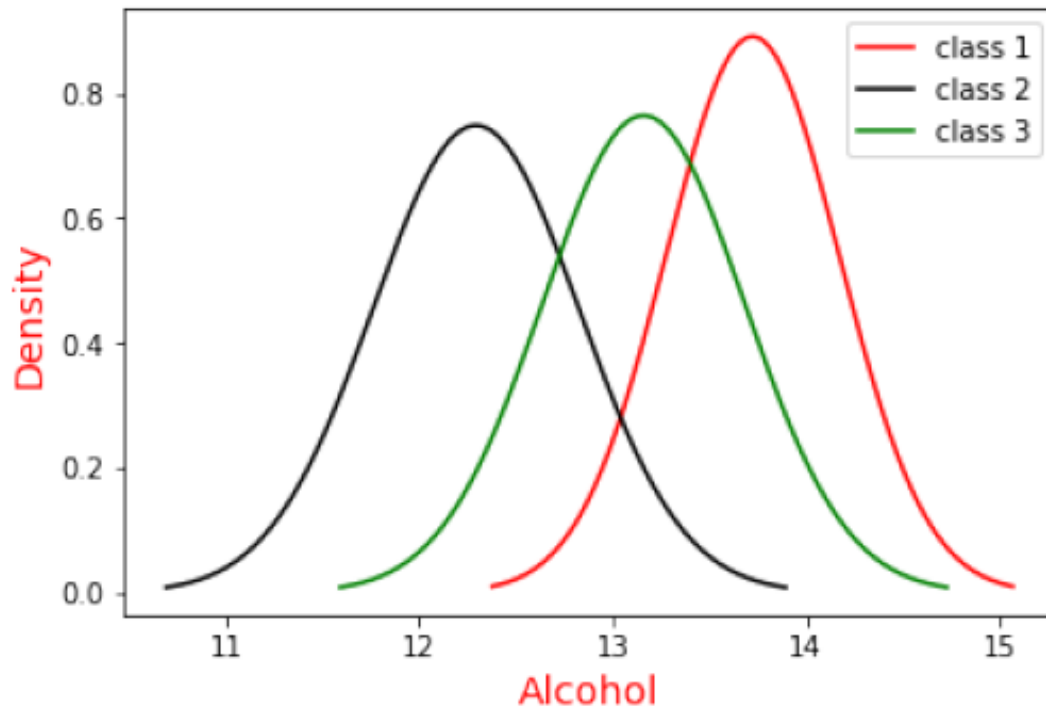

Figure 1: Estimate of confusion matrix for orginal data and normalized data using leave-one-out cross-validation (LOOCV) and 1-NN classification with Euclidean distance.

(a) x has mean 2 and standard deviation 1, y has mean 2 and standard deviation 0.5, and the correlation between x and y is -0.5.

(b) x has mean 1 and standard deviation 1, and y is equal to 2.