DSC 140A - Homework 08

Due: Wednesday, March 6

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. Unless otherwise noted by the problem's instructions, show your work or provide some justification for your answer. Homeworks are due via Gradescope at 11:59 PM.

Problem 1.

The file at the link below contains a data set of 100 points from two classes (1 and -1).

`https://f000.backblazeb2.com/file/jeldridge-data/003-two_clusters/data.csv`

The first two columns contains features, and the last column contains the label of the point. Note that the labels are 1 and -1, not 1 and 0, and that there are no column headers.

In all parts of this problem you may use code to compute your answers. If you do, be sure to include your code.

a) Suppose two Gaussians with full covariance matrices are used to model the densities $p_X(x \mid Y = 1)$ and $p_X(x \mid Y = -1)$. What are the maximum likelihood estimates for the covariance matrices of each Gaussian?

(Allow each Gaussian to have its own covariance matrix; don't use the same covariance for both.)

Hint: the covariance matrix for the Gaussian fit to points from class 1 should have 12.29 in its top-left entry.

Code For Both Parts

```python
import numpy as np

data = np.loadtxt('data.csv', delimiter=',')
X = data[:, :2]
y = data[:, 2]

class_1 = X[y == 1]
class_neg1 = X[y == -1]

mu_1 = np.mean(class_1, axis=0)
mu_neg1 = np.mean(class_neg1, axis=0)

def cov(X, mu):
    n = X.shape[0]
    return (X - mu).T @ (X - mu) / n

cov_1 = cov(class_1, mu_1)
cov_neg1 = cov(class_neg1, mu_neg1)

print(
    f'Mean for class 1: {mu_1}\n'
    f'Mean for class -1: {mu_neg1}\n'
    f'Covariance for class 1: \n{cov_1}\n'
    f'Covariance for class -1:\n {cov_neg1}'
)
```

```python
x_news = np.array([
    [0, 0],
    [1, 1],
    [10, 5],
    [5, -5],
    [8, 5]
])

def predict(x_new, mu_1, mu_2, cov_1, cov_2):
    d_1 = (x_new - mu_1) @ np.linalg.inv(cov_1) @ (x_new - mu_1).T
    d_2 = (x_new - mu_2) @ np.linalg.inv(cov_2) @ (x_new - mu_2).T
    return 1 if d_1 < d_2 else -1

for x in x_news:
    pred = predict(x, mu_1, mu_neg1, cov_1, cov_neg1)
    print(f'Prediction for {x}: {pred}')
```

Solution: The covariance matrix for the Gaussian fit the maximum likelihood estimates for the covariance matrices of each Gaussian are:

$$cov_1 = \begin{bmatrix} 12.29584016 & 0.28098224 \\ 0.28098224 & 16.06766736 \end{bmatrix}, \quad cov_{-1} = \begin{bmatrix} 10.91736224 & 0.53015728 \\ 0.53015728 & 15.17320916 \end{bmatrix}$$

b) Using the estimated Gaussians with the Bayes classification rule, what are the predicted labels of each of the following points?

- $(0, 0)^T$
- $(1, 1)^T$
- $(10, 5)^T$
- $(5, -5)^T$
- $(8, 5)^T$

Show your calculations.

Note: making predictions in this way (using Gaussians with unequal covariance matrices) is known as *Quadratic Discriminant Analysis*.

Solution: Making the predictions of the new points above we get,

Point	Predicted Label
$(0, 0)^T$	-1
$(1, 1)^T$	-1
$(10, 5)^T$	1
$(5, -5)^T$	-1
$(8, 5)^T$	-1

Problem 2.

In lecture, we derived Linear Discriminant Analysis (LDA) by starting with the Bayes classifier and modeling each class-conditional density as a multivariate Gaussian and using the same covariance matrix for each. We stated, but did not prove, that the decision boundary of an LDA classifier is linear.

Recall that, for a binary classifier based on the Bayes Classifier, the decision boundary is the set of all points $\vec{x}$ where

$$\hat{p}(\vec{x} \mid Y = 1)\hat{\mathbb{P}}(Y = 1) = \hat{p}(\vec{x} \mid Y = 0)\hat{\mathbb{P}}(Y = 0),$$

where the various $\hat{p}$ and $\hat{P}$ are estimated densities and probabilities.

Using this fact, prove that the decision boundary of an LDA classifier is linear. For simplicity, you may assume that $\vec{x} \in \mathbb{R}^2$ and that the shared covariance matrix is diagonal (although the result holds even if the covariance matrix is not diagonal).

Hint: since you may assume that $\vec{x} = (x_1, x_2)^T$, you can start from the above equality and solve for x_2 in terms of x_1, showing that you get the equation of a line.

Solution: Using the fact that we have a shared covariance matrix, we can write the gaussian class-conditional densities as

$$p(\vec{x} \mid Y = y) = \frac{1}{(2\pi)^{d/2}|C|^{1/2}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu}_y)^T C^{-1}(\vec{x} - \vec{\mu}_y)\right)$$

which as the same covariance matrix C for both classes. We can then write the decision boundary as

$$\begin{aligned}
\exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu}_i)^T C^{-1}(\vec{x} - \vec{\mu}_i)\right)\mathbb{P}(Y = 1) &= \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu}_j)^T C^{-1}(\vec{x} - \vec{\mu}_j)\right)\mathbb{P}(Y = 0) \\
-\frac{1}{2\sigma^2}(x - \mu_i)^T(x - \mu_i) + \ln(\mathbb{P}(Y = 1)) &= -\frac{1}{2\sigma^2}(x - \mu_j)^T(x - \mu_j) + \ln(\mathbb{P}(Y = 0)) \\
-\frac{1}{2\sigma^2}(x - \mu_i)^T(x - \mu_i) + \frac{1}{2\sigma^2}(x - \mu_j)^T(x - \mu_j) &= \ln\left(\frac{\mathbb{P}(Y = 0)}{\mathbb{P}(Y = 1)}\right) \\
\frac{1}{2\sigma^2}\left((x - \mu_j)^T(x - \mu_j) - (x - \mu_i)^T(x - \mu_i)\right) &= \ln\left(\frac{\mathbb{P}(Y = 0)}{\mathbb{P}(Y = 1)}\right)
\end{aligned}$$

We can then expand the terms in the parentheses to get

$$(\vec{x} - \mu_i)^T(\vec{x} - \mu_i) = \frac{(x_1 - \mu_{i1})^2}{\sigma^2} + \frac{(x_2 - \mu_{i2})^2}{\sigma^2}$$

So we can write the decision boundary as

$$\begin{aligned}
(x_1 - \mu_{j1})^2 + (x_2 - \mu_{j2})^2 - (x_1 - \mu_{i1})^2 - (x_2 - \mu_{i2})^2 &= 2\sigma^2 \ln\left(\frac{\mathbb{P}(Y = 0)}{\mathbb{P}(Y = 1)}\right) \\
x_1^2 - 2x_1\mu_{j1} + \mu_{j1}^2 + x_2^2 - 2x_2\mu_{j2} + \mu_{j2}^2 - x_1^2 + 2x_1\mu_{i1} - \mu_{i1}^2 - x_2^2 + 2x_2\mu_{i2} - \mu_{i2}^2 &= 2\sigma^2 \ln\left(\frac{\mathbb{P}(Y = 0)}{\mathbb{P}(Y = 1)}\right) \\
-2x_1\mu_{j1} + 2x_1\mu_{i1} - 2x_2\mu_{j2} + 2x_2\mu_{i2} &= 2\sigma^2 \ln\left(\frac{\mathbb{P}(Y = 0)}{\mathbb{P}(Y = 1)}\right) - \mu_{j1}^2 - \mu_{j2}^2 + \mu_{i1}^2 + \mu_{i2}^2
\end{aligned}$$

So in the context of a linear equation $Ax_1 + Bx_2 + C = 0$, we can collect the like terms to get

$$A = 2(\mu_{i1} - \mu_{j1}), \quad B = 2(\mu_{i2} - \mu_{j2}), \quad C = 2\sigma^2 \ln\left(\frac{\mathbb{P}(Y = 0)}{\mathbb{P}(Y = 1)}\right) - \mu_{j1}^2 + \mu_{j2}^2 - \mu_{i1}^2 + \mu_{i2}^2$$

Where we know that $\sigma, \mathbb{P}(Y = 0), \mathbb{P}(Y = 1), \mu_{i1}, \mu_{i2}, \mu_{j1}, \mu_{j2}$ are all constants, so we can conclude that the decision boundary is linear.

Problem 3.

You've been hired by a generic online retailer named after a rainforest named after a river. Your job is to build a model to predict whether or not a particular item will sell. You are provided with a dataset of outcomes for a collection of products:

Brand	Price Range	Condition	Sold
A	High	Used	No
A	High	New	Yes
B	Low	New	Yes
C	Medium	New	Yes
B	Low	Used	No
A	High	New	No
C	High	Used	Yes
A	Medium	Used	Yes
B	Medium	Used	No
C	Low	New	No
B	Low	Used	Yes

Using a Naïve Bayes classifier and the data above, predict if a product with Brand = B, Price Range = Medium, Condition = Used will sell or not. Show your calculations.

Solution: First we get the probabilities of each the class conditional probabilities, for the class y =Yes and y =No.

$$P(\text{Brand} = B|y = \text{Yes}) = \frac{2}{6}$$
$$P(\text{Brand} = B|y = \text{No}) = \frac{2}{5}$$
$$P(\text{Price Range} = \text{Medium}|y = \text{Yes}) = \frac{2}{6}$$
$$P(\text{Price Range} = \text{Medium}|y = \text{No}) = \frac{1}{5}$$
$$P(\text{Condition} = \text{Used}|y = \text{Yes}) = \frac{3}{6}$$
$$P(\text{Condition} = \text{Used}|y = \text{No}) = \frac{3}{5}$$

Then next we calculate the prior probabilities of each class given the conditions that we have.

$$\begin{aligned} &P(\text{Brand} = B \text{ and Price Range} = \text{Medium and Condition} = \text{Used}|y = \text{Yes}) \\ &= P(\text{Brand} = B|y = \text{Yes}) \cdot P(\text{Price Range} = \text{Medium}|y = \text{Yes}) \cdot P(\text{Condition} = \text{Used}|y = \text{Yes}) \\ &= \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{18} \end{aligned}$$

For the No class we have

$$\begin{aligned} &P(\text{Brand} = B \text{ and Price Range} = \text{Medium and Condition} = \text{Used}|y = \text{No}) \\ &= P(\text{Brand} = B|y = \text{No}) \cdot P(\text{Price Range} = \text{Medium}|y = \text{No}) \cdot P(\text{Condition} = \text{Used}|y = \text{No}) \\ &= \frac{2}{5} \cdot \frac{1}{5} \cdot \frac{3}{5} = \frac{6}{125} \end{aligned}$$

Using the Bayes rule we can calculate the probability of the class given the conditions that we have

$$\begin{aligned}&P(\text{Brand} = B \text{ and Price Range} = \text{Medium and Condition} = \text{Used}|y = \text{No})P(y = \text{No})\\&\quad= \frac{6}{125} \cdot \frac{5}{11} \approx 0.0218\end{aligned}$$

and

$$\begin{aligned}&P(\text{Brand} = B \text{ and Price Range} = \text{Medium and Condition} = \text{Used}|y = \text{Yes})P(y = \text{Yes})\\&\quad= \frac{1}{18} \cdot \frac{6}{11} =\approx 0.0303\end{aligned}$$

As $0.0303 > 0.0218$ we can conclude that the product will sell.