ECE/CS/ME 539 – Fall 2024 — Homework 3

Problem 1: Confusion Matrix (3 points)

A new test is developed to detect spam emails. After an experiment, the confusion matrix of this test is reported below:

Actual / Predicted	**P**	**N**
P	15	5
N	4	16

- How many spam (P) results does this new test report? _____
- What percentage of actual spam emails are correctly identified as spam in this test? _____
- What is the false positive rate, defined as the fraction of products that are reported as defective but are actually non-defective, among all negative tests? _____

Problem 2: Performance Metrics (4 points)

Consider 10 feature-label pairs $(x(k), y(k); 1 \leq k \leq 10)$. Assume that $y(k) \in \{0, 1\}$. The posterior probability $P(y(k) = 1|x(k))$ is given in the 2nd row of the table below, and the corresponding ground truth label is given in the third row. Here, a class label = 0 means Negative, and a class label = 1 means Positive.

Index (k)	1	2	3	4	5	6	7	8	9	10
$P(y(k) = 1\|x(k))$	0.05	0.15	0.40	0.55	0.25	0.45	0.48	0.62	0.67	0.75
True Label	0	0	0	0	1	1	1	1	1	1
Predicted label $y(k)$										

Given a threshold b, we set the predicted label $y(k) = 0$ if $P(y(k) = 1|x(k)) \leq b$; and $= 1$ otherwise.

(a) If $b = 0.3$, fill in the predicted label in the 4th row of the above table.

(b) Compute the confusion matrix C with $b = 0.3$.

(c) With $b = 0.3$, compute the following quantities: sensitivity (sen), specificity (spe), Pr. False Alarm (pfa), Pr. Miss (pmiss), precision (pre), recall, and accuracy.

(d) For the value of threshold b varying from 0 to 1, compute the list of distinct pairs of (TPR, FPR) and then plot the ROC curve and calculate the area under the ROC curve (AUC).

Problem 3: PCA (5 points)

`MNIST` consists of 28 by 28 gray level images of hand-written numerals from 0 to 9. You can download the dataset using the code below:

```
import torchvision

data_transform = torchvision.transforms.Compose([
    torchvision.transforms.ToTensor(),
    lambda x: torch.floor(x * 255 / 128).squeeze(dim=0)
])
mnist_test = torchvision.datasets.MNIST(
root='./temp', train=False, transform=data_transform, download=True)
x, y = mnist_test.__getitem__(1010) # x is the image and y is the
    corresponding label.
```

(a) Number of samples $N =$ _____. Feature dimension = _____.

(b) Visualize the first 20 rows (samples). Each should be displayed as a 28 by 28 image. Refer to the `d2l 22.9`.

(c) Denote the $N \times d$ feature matrix as X. Perform SVD of X. Design the singular values as a vector s. Plot $\log_{10}(s)$ over the range 1 to d.

(d) Denote the first two principal components by a $d \times 2$ matrix V. Use the first 2 principal components, projecting each row of the X matrix by computing $Z = XV$. Each row of Z is a 1×2 vector corresponding to a point in a 2D space spanned by the two columns of V. Give a scatter plot of these projected 2D points corresponding to numerals 0 and 9. Note that the numerals are class labels.

(e) This one is for all 10 numerals. An approximation of the original feature matrix X may be estimated as:

$$\hat{X} = XVV^T = ZV^T$$

Visualize the corresponding 28 by 28 patterns of the first 20 rows of $\hat{X}$