CSE 151A - Homework 02

Due: Wednesday, April 15, 2020

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 on Wednesday at 11:59 p.m.

Essential Problem 1.

The table below shows the SAT scores for 10 randomly-selected applicants to UCSD, along with whether or not they were admitted.

Score	Admitted?
1240	Y
1310	Y
1470	Y
1500	Y
1200	N
1200	N
1220	N
1250	N
1290	N
1400	N

Your friend is applying to UCSD with an SAT score of 1300. Using Gaussians to estimate the conditional probabilities involved, use a Bayes classifier to predict whether your friend will be admitted or not. Show your work and all calculations involved.

Hint: You can use `scipy.stats.norm.pdf` to from the Python package `scipy` to evaluate the normal PDF (or another similar function in a different language). But make sure you know how the function works. In particular, does it require the standard deviation, or the variance?

Solution: The data is split into two classes: Y and N. We fit a Gaussian to each by computing the mean and variance of each group of data independently. The admitted students have a mean score of 1380, with a standard deviation of approximately 125. The non-admitted students have a mean score of 1260 with a standard deviation of approximately 76. There are 4 admitted students and 6 non-admitted students.

Note: There are two "versions" of the standard deviation: one in which we divide by $1/(n-1)$, and another in which we divide by $1/n$. In statistical terms, the first is the unbiased estimator and as such some people prefer it. However, the difference between them is minuscule when n is large. The standard deviations above were calculated using `pandas`, which uses $1/(n-1)$. If we were to calculate them with `numpy`'s `np.std` instead, we would find the standard deviation of admitted students' scores to be around 108, and the standard deviation of non-admitted scores to be about 70.

We calculate:

$$P(\text{Score} = 1300 \,|\, \text{Admitted? } = \text{Y})P(\text{Admitted? } = \text{Y})$$

and

$$P(\text{Score} = 1300 \,|\, \text{Admitted? } = \text{N})P(\text{Admitted? } = \text{N})$$

Right off the bat, we know $P(\text{Admitted? } = \text{Y}) = 0.4$, while $P(\text{Admitted? } = \text{N}) = 0.6$. Next we estimate $P(\text{Score} = 1300 \,|\, \text{Admitted? } = \text{Y})$ using a Gaussian with mean 1380 and standard deviation 125. We

could do this with `scipy`, for instance:

```
scipy.stats.norm.pdf(1300, 1380, 125)
```

This evaluates to 0.0026. Next, for the second Gaussian, we compute:

```
scipy.stats.norm.pdf(1300, 1260, 76)
```

This evaluates to 0.0045. Therefore

$$P(\text{Score} = 1300 \,|\, \text{Admitted? = Y})P(\text{Admitted? = Y}) = (0.0026)(0.4) = 0.001$$

and

$$P(\text{Score} = 1300 \,|\, \text{Admitted? = N})P(\text{Admitted? = N}) = (0.0045)(0.6) = 0.0027$$

As a result, we predict that your friend will not be accepted. In fact, our model predicts that it is almost three times more likely that they will be rejected than accepted.

Note: If we use $1/n$ instead of $1/(n-1)$ when calculating the standard deviation, we would find

$$P(\text{Score} = 1300 \,|\, \text{Admitted? = Y})P(\text{Admitted? = Y}) = (0.0028)(0.4) = 0.001$$

and

$$P(\text{Score} = 1300 \,|\, \text{Admitted? = N})P(\text{Admitted? = N}) = (0.0048)(0.6) = 0.0029$$

Essential Problem 2.

The CDC is testing a vaccine for COVID-19 and has gathered a group of 44 people to experiment on. The people are classified by the state they are from, as shown in the table below.

State	#
Ohio	12
California	27
Texas	5

Suppose that 3 of the people from Ohio have COVID-19, 10 of the people from California have the virus, and one person from Texas has the virus. You randomly select one person from the study and learn that they are healthy – they do not have the virus. What is the probability that the person is from Texas?

Solution:

We want $P(\text{Texas} \,|\, \text{Healthy})$. We will compute this with Bayes' Theorem:

$$\begin{aligned} P(\text{Texas} \,|\, \text{Healthy}) &= \frac{P(\text{Healthy} \,|\, \text{Texas}) \cdot P(\text{Texas})}{P(\text{Healthy})} \\ &= \frac{4/5 \cdot 5/44}{30/44} \\ &= 2/15 \approx 1.33 \end{aligned}$$

Here we used the fact that the probability of random person being healthy is $30/44$, since the probability that a random person is sick is $(3 + 10 + 1)/44 = 14/44$.

Essential Problem 3.

In each part below, assume that you have gathered a data set consisting of the quantities described. Respond with a matrix containing the *sign* of each entry of the data's covariance matrix.

In each case there will be a preferred answer – but the correct answer isn't necessarily unique. If you feel

unsure as to whether the sign of the covariance between two variables is positive or negative, make a guess and provide your reasoning. Otherwise, you do not need to show your work for this problem if you don't want to.

Example: Let X_1 be a person's height, and X_2 be their weight.

Solution: The signs of the entries of the covariance matrix are $\begin{pmatrix} + & + \\ + & + \end{pmatrix}$ because a person's weight tends to be larger the taller they are.

a) Let X_1 be a person's midterm score, let X_2 be their final exam score, and let X_3 be their GPA.

Solution: Remember that the covariance of random variables X and Y is given by

$$\mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$$

So if when X is above its average, Y also tends to be above its average, then the covariance is positive. On the other hand, if when X is above average, Y is *below* its average, the covariance will be negative.

The covariance of a variable with itself is called the variance, and it is always positive.

In this case, if a person scores above average on their first midterm, they are likely to score above average on their final exam, so the covariance between X_1 and X_2 is probably positive. Likewise, if a person scores better than average on their midterm, they are likely to have a higher than average GPA, so the covariance of X_2 and X_3 is positive. Likewise, the covariance of X_2 and X_3 is likely positive.

As a result, the signs of the entries of the covariance matrix are:

$$\begin{pmatrix} + & + & + \\ + & + & + \\ + & + & + \end{pmatrix}$$

b) For a particular day, let let X_1 be the temperature, X_2 be the number of hours that the air conditioner ran, and X_3 be the number of winter coats sold on that day.

Solution: As temperature increases, so does air conditioner usage, so the covariance of X_1 and X_2 is positive. But as the temperature increases, sales of winter coats decrease; the covariance of X_1 and X_3 is negative. Likewise, as air conditioner usage increases, sales of winter coats probably decrease too (since the air condition usage suggests that it is hot outside).

Therefore the signs of the entries of the covariance matrix are:

$$\begin{pmatrix} + & + & - \\ + & + & - \\ - & - & + \end{pmatrix}$$

c) Let X_1 be the longest distance a person can run, let X_2 be their age, and let X_3 be a measure of the efficiency of their lungs (the larger X_3, the more efficient their lungs).

Solution: As a person's age increases, the distance they can run decreases (assuming that they're an adult). So the covariance of X_1 and X_2 is negative. As their age increases, the efficiency of their lungs decreases as well, so the covariance of X_2 and X_3 is negative. And as someone is able to run longer distances, the efficiency of their lungs increases, so the covariance of X_1 and X_3 is

positive.

$$\begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}$$

Essential Problem 4.

Let X be an $n \times d$ matrix, let A be a $d \times r$ matrix, and let B be an $r \times n$ matrix. Let $\vec{x}$ be a vector in $\mathbb{R}^n$ (that is, an $n \times 1$ column vector), let $\vec{y}$ be a vector in $\vec{R}^d$ (that is, a $d \times 1$ column vector). For each of the following, state whether the result is a scalar, a vector, or a matrix. If it is a vector or a matrix, state its shape (number of rows and columns).

For the purposes of this question, a matrix with one column is considered a column vector, and a matrix with one row is considered a row vector. If the result of an expression is 1×1, it is a scalar. You do not need to show your work.

a) $\vec{x} \cdot \vec{x}$

Solution: Scalar.

b) XA

Solution: $n \times r$ matrix.

c) $XX^\intercal$

Solution: $n \times n$ matrix

d) $X^\intercal X$

Solution: $d \times d$ matrix

e) $(XA)^\intercal \vec{x}$

Solution: $r \times 1$ vector.

f) $\vec{y}^\intercal \vec{y} (XX^\intercal)^{-1}$

Solution: $n \times n$ matrix.

g) $(\vec{x} \cdot \vec{x} + \vec{y} \cdot \vec{y}) + x^\intercal B^\intercal A^\intercal X^\intercal X A B \vec{x}$

Solution: Scalar.

h) $B^\intercal A^\intercal X^\intercal X A B$

Solution: $n \times n$ matrix.

Essential Problem 5.

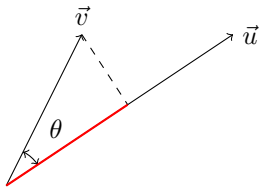
We will soon need to remember the key properties of the dot product. This question is meant to help you remember them.

a) Recall from your class on vector algebra that one way to define the dot product of two vectors, $\vec{u}$ and $\vec{v}$, is:

$$\vec{u} \cdot \vec{v} = \|\vec{u}\|\|\vec{v}\| \cos\theta,$$

where $\|\vec{u}\|$ is the length of the vector $\vec{u}$, $\|\vec{v}\|$ is the length of $\vec{v}$, and θ is the angle between the two vectors.

Two vectors $\vec{u}$ and $\vec{v}$ are shown below.



Argue that the length of the red segment is $(\vec{u} \cdot \vec{v})/\|\vec{u}\|$.

Solution: We see a right triangle whose hypotenuse has length $\|\vec{v}\|$. The length of the adjacent (red) side is given by the length of the hypotenuse by the cosine of θ:

$$\|\vec{v}\| \cos\theta$$

Since $\vec{u} \cdot \vec{v} = \|\vec{u}\|\|\vec{v}\| \cos\theta$, the length of the red segment must be $(\vec{u} \cdot \vec{v})/\|\vec{u}\|$.

b) The function $f(\vec{x}) = 5x_1 + 2x_2 - 3x_3$ can be written as $f(\vec{x}) = \vec{w} \cdot \vec{x}$ for some vector $\vec{w}$. What is $\vec{w}$?

Solution: $\vec{w} = (5, 2, -3)^T$

c) Let $\vec{x} = (x_1, \ldots, x_d)^T$ be a vector in $\mathbb{R}^d$. If $\vec{x}$ is a unit vector (that is, the length of $\vec{x}$ is 1) and $x_1 = 0.1$, what is the largest that any of the remaining entries $x_2, \ldots, x_d$ can possibly be?

Solution: We solve $(0.1)^2 + x^2 = 1$ for x and find that $x = \sqrt{1 - .01} = \sqrt{0.99}$.

d) What is the angle between $\vec{x} = (1, 2, 3)^T$ and $\vec{y} = (3, 2, 1)^T$?

Solution: We solve $\vec{x} \cdot \vec{y} = \|\vec{x}\|\|\vec{y}\| \cos\theta$ for θ, and find that $\theta = \arccos(\frac{10}{14}) = 0.775 = 44.4°$.

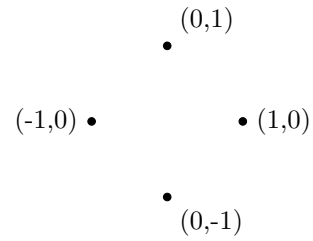
Plus Problem 1. (5 plus points)

In this problem, let X_1 and X_2 be random variables.

a) Show that if X_1 and X_2 are independent, then $\text{Cov}(X_1, X_2) = 0$

b) Independence implies zero covariance, but zero covariance does not imply independence in general. Here's an example demonstrating this.

Consider the four points below:



A point is chosen from these four, uniformly at random. Let X be its x-coordinate, and let Y be its y-coordinate. Show that X and Y are dependent, but that $\text{Cov}(X, Y) = 0$.

Solution: First, X and Y are not independent. For instance,

$$P(X = 1, Y = 0) = 1/4$$

however this is not equal to $P(X = 1)P(Y = 0)$, since $P(X = 1) = 1/4$ and $P(Y = 0) = 1/2$.

Now we compute the covariance.

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$$

Observe that $\mathbb{E}X = 0$ and $\mathbb{E}Y = 0$. So:

$$\begin{aligned}
&= \mathbb{E}[XY] \\
&= (1 \cdot 0)P(X = 1, Y = 0) + (0 \cdot 1)P(X = 0, Y = 1) \\
&\quad + (-1 \cdot 0)P(X = -1, Y = 0) + (0 \cdot -1)P(X = 0, Y = -1) \\
&= 0
\end{aligned}$$

c) Zero covariance does not imply independence in general. However, in the special case that X and Y are *jointly* Gaussian random variables, $\text{Cov}(X, Y) = 0$ *does* imply that X and Y are independent. Recall that random variables X and Y are *jointly* Gaussian if the random vector $\vec{S} = (X, Y)^T$ has a density which is a two-dimensional Gaussian. The statement that X and Y have zero covariance is saying that the covariance matrix of the Gaussian describing their joint density is diagonal.

Prove that jointly Gaussian random variables with zero covariance are independent.

Solution: If X and Y are jointly Gaussian, then their joint distribution is

$$P(x, y) = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\vec{s} - \vec{\mu})^T \Sigma^{-1} (\vec{s} - \vec{\mu})\right),$$

where $\vec{s} = (x, y)^T$ and $\vec{\mu}$ is the vector $(\mathbb{E}X, \mathbb{E}Y)^T$ and Σ is the covariance matrix. We will show that $P(x, y) = P(x)P(y)$, and so X and Y are independent.

In this case X and Y have zero covariance, so the covariance matrix is diagonal. That is:

$$\Sigma = \begin{pmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{pmatrix},$$

where σ_X^2 and σ_Y^2 are the variances of X and Y, respectively. Then:

$$(\vec{s}-\vec{\mu})^T\Sigma^{-1}(\vec{s}-\vec{\mu}) = [(x,y)-(\mu_X,\mu_Y)]\begin{pmatrix}1/\sigma_X^2 & 0\\ 0 & 1/\sigma_Y^2\end{pmatrix}[(x,y)^T-(\mu_X,\mu_Y)^T]$$
$$= \frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}$$

Similarly, $|\Sigma| = \sigma_X^2\sigma_Y^2$. And so:

$$P(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y}\exp\left(-\left[\frac{(x-\mu_X)^2}{2\sigma_X^2}+\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right]\right),$$
$$= \frac{1}{\sqrt{2\pi}\sigma_X}\exp\left(-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right)+\frac{1}{\sqrt{2\pi}\sigma_Y}\exp\left(-\frac{(x-\mu_Y)^2}{2\sigma_Y^2}\right)$$
$$= P(x)P(y)$$

Plus Problem 2. (9 plus points)

Tumors are often diagnosed as malignant or benign through medical imaging. The file http://cse151a.com/data/cancer/train.csv contains data on 400 tumors collected as part of a breast cancer study at the University of Wisconsin. The data contains 30 measurements of each tumor, including the tumor's area, its perimeter, a measure of its texture, and so on. All features are continuous. The first column of the data reports whether the tumor was benign (`B`) or malignant (`M`). The file http://cse151a.com/data/cancer/test.csv contains a test set.

a) Create scatter plots of the `radius_mean` column versus the `texture_mean` column for benign and malignant tumors using the training data. Overlay your plots on the same graph.

Solution:

See https://go.ucsd.edu/2VxSIrk for a solution notebook.

b) Perform Linear Discriminant Analysis by estimating each class-conditional density with a Gaussian; the two Gaussians should share the same diagonal covariance matrix. Standardize each feature before performing your analysis. Report the error of your classifier on both the training set and the test set. Provide your code.

Hint 0: You can use libraries like `scipy` to evaluate the multivariate Normal pdf, but don't use code which performs LDA itself.

Hint 1: The top left entry of your shared covariance matrix should be roughly 0.46.

Hint 2: How do you get one covariance matrix for both classes? The lecture describes the standard approach.

Hint 3: The test set should be standardized too. When standardizing it, what makes the most sense: using the mean and variance from the training set, or from the test set? Oftentimes in practice we don't see the whole test set at once, but rather see one point at a time – you can make that limiting assumption here.

Solution: Training error: 6%. Test error: 5.9%.

See https://go.ucsd.edu/2VxSIrk for a solution notebook.

c) Perform Quadratic Discriminant Analysis by estimating each class-conditional density with a Gaussian; the two Gaussians should have different full covariance matrices. Standardize each feature before

performing your analysis. Report the error of your classifier on both the training set and the test set. Provide your code.

Solution: Train error: 3.2%. Test error: 3%.

See https://go.ucsd.edu/2VxSIrk for a solution notebook.