HW 0

Nicole Sullivan

1a

> Answer the following questions:
> Which of the following courses have you taken?
> i. Artificial Intelligence II
> ii. Introduction to Data Mining
> iii. (Advanced) Algorithms

I haven't taken any of these.

1b

> Have you taken any course on Probability/Statistics? If yes, please write down the course title (not number) and department.

- Bayesian Astrostatistics, College of Science and Engineering
- Applied Regression Analysis, Department of Statistics (UChicago)
- Data Science, Department of Statistics (UChicago)
- Computing for the Social Sciences, Department of Social Sciences (UChicago)

1c

> Have you taken any course on Numerical Methods/Linear Algebra/Multivar Calculus? If yes, please write down the course title (not number) and department.

- Linear Algebra/Differential Equations, Mathematics Department (MCTC)
- Multivariable Calculus, Mathematics Department (MCTC)

2a

> Given a full rank matrix A ∈ $\mathbb{R}^{m\times n}$ where m > n and B ∈ $\mathbb{R}^{m\times k}$, show how to solve the following system of equations:

$$AX = B$$

First, we can begin with a system of equations in terms of the matrix A:

$$A\vec{x} = \vec{0} \qquad N(A) \qquad \text{complementary solution}$$

$$A\vec{x} = \vec{b} \qquad C(A) \qquad \text{particular solution}$$

Adding the two above together, we get:

$$A\vec{x_n} + A\vec{x_p} = \vec{0} + \vec{b}$$

We use an n subscript to denote the x vector in the complementary solution, since it pertains to the solution across the null space.

Then, since matrix multiplication is distributive, we get:

$$A(\vec{x_n} + \vec{x_p}) = \vec{b}$$

Therefore, the general solution for $Ax = B$ is given by:

$$\vec{x} = \vec{x_n} + \vec{x_p}$$

To solve for the general solution, we would just need to augment the matrix with the $\vec{b}$ (in the case of the particular solution), use Gauss-Jordan elimination to put the matrix in reduced row echelon form, and then use the corresponding system of equations to determine $b_1, b_2, b_3 \ldots b_k$. In the case of the complementary solution, we simply take the matrix in reduced row echelon form and augment it with the zero vector, $\vec{0}$.

2b

> What happens if A is not full rank? Briefly explain your answer.

If A is not full rank, then not all the vectors are linearly independent, and, therefore, the reduced row echelon form of the augmented matrix won't provide a sufficient number of equations to solve for the unknowns.

3

> Given two vectors $\vec{a}$ and $\vec{b}$, explain one way to calculate the distance between the vectors (you can give an example). Show the same for two matrices A and B. Note: your distance should be one scalar number that represents the distance between the two vectors/matrices.

The formula for the Euclidean distance between two vectors is:

$$d(\vec{a},\vec{b}) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \ldots (a_k - b_k)^2}$$

where the subscript indicates the component of the vector being manipulated.

Then, to find the distance between, for example, the vectors $\vec{a} = (1, 1, 2)$ and $\vec{b} = (-1, 3, 0)$, the formula would be:

$$\begin{aligned} d(\vec{a},\vec{b}) &= \sqrt{(1-(-1))^2 + (1-3)^2 + (2-0)^2} \\ &= \sqrt{2^2 + (-2)^2 + 2^2} \\ &= \sqrt{4+4+4} = \sqrt{12} \text{ or } 2\sqrt{3} \end{aligned}$$

To find the distance between 2 matrices, we can use the Frobenius Norm:

$$||C||_F = \sqrt{\mathrm{Tr}(CC^H)}$$

where:

- Tr is the trace of the matrix, or the sum of the diagonal elements
- C is the difference matrix between two matrices, A and B
- C^H is the conjugate transpose of the matrix C

For example, if

$$A = \begin{bmatrix} 4 & 1 \\ 2 & 1 \end{bmatrix}$$

and

$$B = \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix}$$

then

$$A - B = C = \begin{bmatrix} 2 & -3 \\ -1 & 0 \end{bmatrix}$$

and

$$C^T = \begin{bmatrix} 2 & -1 \\ -3 & 0 \end{bmatrix}$$

In this example, then,

$$CC^T = \begin{bmatrix} 13 & -2 \\ -2 & 1 \end{bmatrix}$$

So $Tr(CC^T) = 14$ and the distance between matrices A and B, then, is $\sqrt{14}$.

4a

> Let D be a random variable for a given disease, assume that the probability a person has the disease is 0.1. Based on this information, researchers developed a new method to say if a person has the disease: for each 10 people that do the test, they randomly report that 1 of them has the disease. Will the method correctly identify if the person has the disease? Briefly explain your answer.

This depends on **conditional probabilities**. If all 10 people were randomly selected to do the test, then the probability that one of the 10 people has the disease is 0.1, and the researchers will fail in about 90% of their reports to identify the correct person. *However*, testing for a disease is *rarely* carried out indiscriminately. Instead, patients are usually screened because they meet certain risk criteria: they display symptoms for a disease, or in the case of screenings like pap smears or mammograms, because of the patient's demographics, family history and/or recency of their last screening. In these cases, the probability a person has the disease is usually higher than the prevalence in the general population (i.e. P(D = 1 | Symptoms = 1) or P(D = 1 | Gender = F & Age = 37)). Therefore, if the researchers are conducting a test amongst a population that has some reason to believe they might have the disease (symptoms, demographics, genetics, etc.), the probability they will correctly identify positives by assigning one positive randomly would be higher than 0.1, and would depend on the relevant conditional probability.

4b

> Another group of researchers developed a new blood test to identify the same disease. The test result is given by a random variable X, with sensitivity and specificity given by 0.7 and 0.8, respectively (that means p(X = 1|D = 1) = 0.7 and p(X = 0|D = 0) = 0.8). If a patient did the blood test and the result is positive, what is the probability that the person has the disease? Hint: you might want to use the Bayes Rule: p(b|a) = p(a|b)p(b)/p(a)

From **4a**, we know that $P(D = 1)$ = 0.1. Putting that together with the new information in part b, we have the following knowns:

- $P(D = 1)$ = 0.1
- $P(X = 1|D = 1)$ = 0.7
- $P(X = 0|D = 0)$ = 0.8

Then, using Bayes' Theorem, we have:

$$P(D = 1|X = 1) = \frac{P(X = 1|D = 1)P(D = 1)}{P(X = 1)}$$

$$P(D = 1|X = 1) = \frac{0.7 \times 0.1}{P(X = 1)}$$

We still have two unknowns, and only one equation. To determine $P(X = 1)$, we'll need to fill in the joint probability table:

	X = 1	**X = 0**
D = 1		
D = 0		

Let's start with filling in what we know from the exercise. Note that, by virtue of knowing $P(D = 1) = 0.1$, we know that $P(D = 0)$, which is just $1 - P(D = 1)$ or 0.9:

	X = 1	**X = 0**	
D = 1			0.1
D = 0			0.9

We have two marginals, and no joint probabilities yet. We can derive $P(X = 0 \cap D = 0)$ first by using the definition of conditional probability:

$$P(X = 0|D = 0) = \frac{P(X = 0 \cap D = 0)}{P(D = 0)}$$

Plugging in our knowns, we get:

$$0.8 = \frac{P(X = 0 \cap D = 0)}{0.9}$$

And rearranging to get our unknown on the left we get:

$$P(X = 0 \cap D = 0) = 0.8 \times 0.9 = 0.72$$

Let's add it to the joint probability table.

	X = 1	**X = 0**	
D = 1			0.1
D = 0		0.72	0.9

And since we have the D = 0 marginal, we can fill in $P(X = 1 \cap D = 0)$ now as well:

	X = 1	**X = 0**	
D = 1			0.1
D = 0	0.18	0.72	0.9

Let's take a second to make sure these probabilities make sense so far. If most of the population *doesn't* have the disease, we would expect that most people also wouldn't have tested positive for it (X = 0) - and we see that that's the case. Great! Let's continue filling in the table.

Next we'll use the conditional probability $P(X = 1|D = 1)$:

$$P(X = 1|D = 1) = \frac{P(X = 1 \cap D = 1)}{P(D = 1)}$$

Again, filling in our knowns we get:

$$0.7 = \frac{P(X = 1 \cap D = 1)}{0.1}$$

And rearranging, we can solve for the intersection:

$$P(X = 1 \cap D = 1) = 0.7 \times 0.1 = 0.07$$

Adding it to our joint probability table, we are nearly finished now:

	X = 1	**X = 0**	
D = 1	0.07		0.1
D = 0	0.18	0.72	0.9

The remaining joint and marginals are just a matter of algebra:

	X = 1	**X = 0**	
D = 1	0.07	0.03	0.1
D = 0	0.18	0.72	0.9
	0.25	0.75	

Now we have our $P(X = 1)$ marginal, which is 0.25. Plugging this into our original set-up, we can now solve for the probability someone has the disease, given their test is positive:

$$P(D = 1|X = 1) = \frac{0.7 \times 0.1}{0.25} = 0.28$$

Or there's a 28% probability that someone has the disease, given that they've tested positive. This seems inordinately low, but we also have to remember that sensitivity and specificity were 0.7 and 0.8, respectively, which are not overly high for an assay, unfortunately.

Just as a thought exercise, what would the probability be if the assay's sensitivity and specificity were both 98%? In that case, our joint probability table would look like this:

	X = 1	**X = 0**	
D = 1	0.098	0.002	0.1
D = 0	0.018	0.882	0.9
	0.116	0.884	

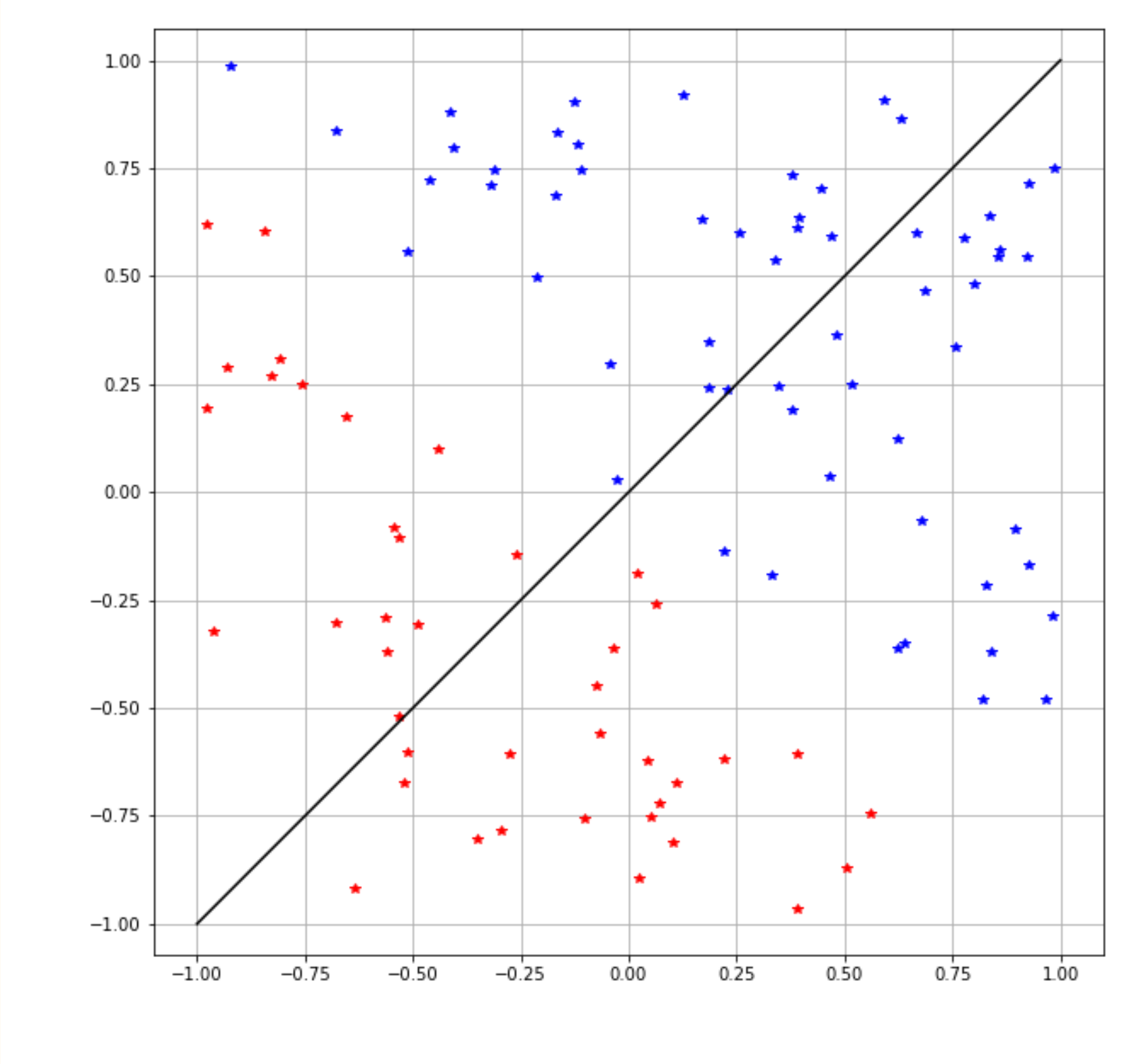
And the probability that someone would have the disease, given a positive test would be $\frac{0.98\times0.1}{0.116} = 0.84$ or there would be an 84% chance the person would have the disease, given a positive test. This illustrates how important an assay's specificity and sensitivity are in ensuring accurate diagnosis!

5

The full solution to question 5 is contained in the `MyPerceptron.py` file.

The plots resulting from running that file are:

Before iteration



After convergence