ECE 760 - Homework 1

Jed Pulley

Problem 1.1

From **Topic 2, Definition 2.3** we can see that a subset $U \subseteq \mathbb{R}^D$ is a *subspace* if for every $a, b \in \mathbb{R}$ and every **u, v** $\in U$, $a\mathbf{u} + b\mathbf{v} \in U$. Stated another way, a subspace must meet the three requirements below:

1. **Closed under addition:** Let $\mathbf{x} = (x_1, x_2, \ldots, x_D)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_D)$ be vectors in $\mathbb{R}^D$. Then their sum $\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_D + y_D)$ is also in $\mathbb{R}^D$. This is because each component of $\mathbf{x} + \mathbf{y}$ is the sum of two real numbers, which are also real numbers.

2. **Closed under scalar multiplication:** Let $\mathbf{x} = (x_1, x_2, \ldots, x_D)$ be a vector in $\mathbb{R}^D$ and let c be a scalar. Then the scalar multiple $c\mathbf{x} = (cx_1, cx_2, \ldots, cx_D)$ is also in $\mathbb{R}^D$. This is because each component of $c\mathbf{x}$ is the product of a real number and a scalar, which is also a real number.

3. **Contains the zero vector:** The vector $\mathbf{0} = (0, 0, \ldots, 0)$ is in $\mathbb{R}^D$, so $\mathbb{R}^D$ is nonempty.

Problem 1.2

(a) To show this, say we have a vector $\mathbf{x}$ in $\mathbb{R}^2$

$$\mathbf{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{1}$$

If we take element-wise square roots of this vector, we get $\sqrt{1}$ equal to 1 and $\sqrt{-1}$ being an imaginary number. Thus, we can show that $\mathbb{R}^D$ is NOT closed under element-wise square roots

(b) An example of a subspace that is closed uner element-wise square roots is $\mathbb{C}^D$

Problem 1.3

As stated in Problem 1.1, to show that something is a subspace, we need to satisfy three properties:

1. **Closed under addition:** Since our in $\mathbb{R}^D$, then we know that any sort of linear operation will remain in the real number space, thus it's closed under addition
2. **Closed under scalar multiplication:** The same above applies for scalar multiplication. So long as we scale by a real number, we'll stay in $\mathbb{R}^D$
3. **Contains the zero vector:** And finally, since our vectors belong to $\mathbb{R}^D$, we know it contains the zero vector.

Thus we can say that $U = span[\mathbf{u}_1, ..., \mathbf{u}_R]$ is a subspace

Problem 1.4

Let's say that the probability that the genes are inactive is represented as

$$\mathbb{P}(A|B) = 0.95$$

and that the probability that you have diabetes is represented as

$$\mathbb{P}(B) = 0.093$$

then using Bayes Rule, we can see that

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}$$

and filling the known values, we have

$$\mathbb{P}(B|A) = \frac{0.95 * 0.093}{\mathbb{P}(A)}$$

(a) An answer cannot be determined given the information.

(b) We would need to know $\mathbb{P}(A)$ which is the percentage of the population has those genes inactive.

(c) Let's just assume that $\mathbb{P}(A) = 0.1$. Meaning that 10% of the population have those genes inactive, resulting in $\mathbb{P}(B|A) = 0.88$, or, there being an 88% chance that you have diabetes given you have those genes inactive. Plainly stated, if the inactive genes are exceedingly uncommon among the population without diabetes, I would be very concerned. 88% is pretty high, I'd be talking to my doctor.

Problem 1.5

Considering the observation that larger delays are rarer than shorter ones, the model I chose is the probability density function of the exponential distribution:

$$t - t_0 \sim \frac{1}{\theta} exp(-\frac{x}{\theta})$$

Here:

- x is our random variable denoting the delay.
- θ is a positive parameter governing the expected delay characteristics, where the larger the value, the longer the delay.

The exponential distribution is a suitable model since it is right-skewed. This aligns with the observation that larger delays are rarer than shorter ones.

Problem 1.6

(a) As shown in in Figure 1 below, I created a small python script using np.random.uniform(0,1,N=1000). Given the randomness, I would say that it is uniform**-ish**, although it definitely could be better.

(b) We can rewrite this as:

$$\mathbb{P}(y_i = 1) = \mathbb{P}(x_i \leq p) = p$$

$$\mathbb{P}(y_i = 0) = \mathbb{P}(x_i > p) = 1 - p$$

which shows that y_i is a Bernoulli distribution where p is our threshold, 1 is the probability of success, and 0 is the probability of failure.

(c) See figures 2, 3, and above. Yes, they match the expected threshold

(d) If z_k is the sum of a batch of y_i's that are themselves of a Bernoulli distribution, then z_k would follow a Binomial distribution. More formally:

$$z_k \sim \mathbb{P}(x = \mathrm{x}) = \binom{n}{\mathrm{x}} p^{\mathrm{x}}(1-p)^{n-\mathrm{x}}, \mathrm{x} = 0, \ldots, n$$

(e) Using np.random.binomial, it shows that they match (d) fairy closely

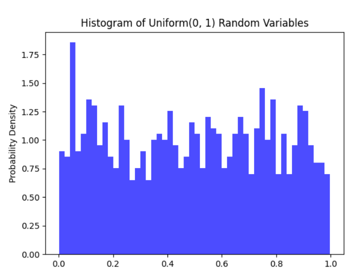

Figure 1: Unfirom Distribution

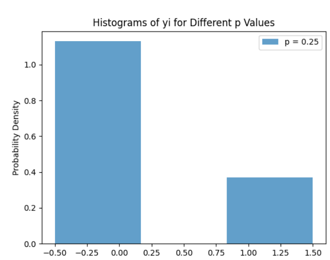

Figure 2: p=0.25

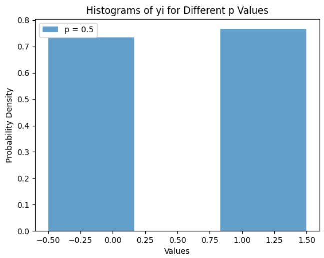

Figure 3: p=0.5

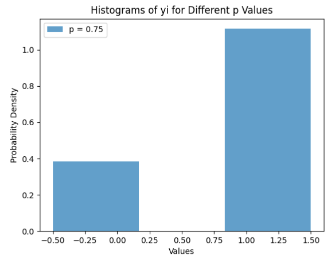

Figure 4: p=0.75

Problem 1.7

With our regresion model like so:

$$l(\theta) = \sum_{\mathrm{i}=1}^{N} [\mathrm{y_i} log(\frac{1}{1+e^{-\theta^T \mathrm{x_i}}}) + (1-\mathrm{y}) log(\frac{1}{1+e^{\theta^T \mathrm{x_i}}})]$$

(a) We can express our gradient like so:

$$\nabla \mathrm{l}(\theta) = \vec{\mathrm{x}}[\sum_{\mathrm{i}=1} n(\frac{1-e^{\theta^T \mathrm{x}}}{1+e^{\theta^T \mathrm{x}}}\mathrm{y_i} + \frac{e^{\theta^T \mathrm{x}}}{1+e^{\theta^T \mathrm{x}}})]$$

(b) And our Hessian will like so:

$$H(l(\theta)) = \tilde{\mathrm{x}}\tilde{\mathrm{x}}^{\mathrm{T}} \sum_{k=1}^{n} \frac{(1-2\mathrm{y_k})e^{\theta^T \mathrm{x}}}{(1+e^{\theta^T \mathrm{x}})^2}$$

(c) As such, $l(\theta)$ is a scalar, $\nabla \mathrm{l}(\theta)$ is a vector, and $H(\mathrm{l}(\theta))$ is a matrix