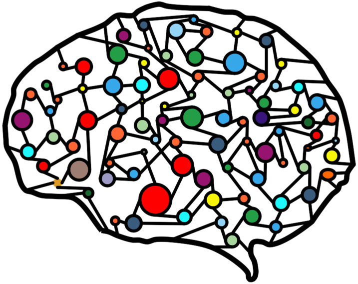
Lecture 28 – Review problems



DSC 40A, Fall 2022 @ UC San Diego

Dr. Truong Son Hy, with help from **many others**

Announcements

- Final exam **tomorrow**!
- Review the solutions to previous homeworks and groupworks.
- Identify which concepts are still iffy. Re-watch lecture and ask questions (now!).
- Look at the past exams at `https://dsc40a.com/resources`.
- Study in groups.
- Make a "cheat sheet".
- Bring a calculator.
- Remember to submit The Course and Professor Evaluations (CAPE) – deadline tonight. If everyone submits CAPE, everyone will get a bonus percentage!

Final schedule

O	40A	Theor Fndtns of Data Sci I (4 Units)						Prerequisites | Resources | Evaluations			
			LE	A00	MWF	3:00p-3:50p	PCYNH	122	Hy, Truong Son		
		88107	DI	A01	M	5:00p-5:50p	PCYNH	122	Hy, Truong Son	9	115
			FI	12/03/2022	S	7:00p-9:59p	CSB	001			
O	40A	Theor Fndtns of Data Sci I (4 Units)						Prerequisites | Resources | Evaluations			
			LE	B00	MWF	4:00p-4:50p	PCYNH	122	Soleymani, Mahdi		
		88109	DI	B01	M	6:00p-6:50p	PCYNH	122	Soleymani, Mahdi	19	115
			FI	12/03/2022	S	7:00p-9:59p	CSB	002			

Time: December 3rd, 2022 – 7:00pm to 10:00pm (3 hours)
Location: CSB building – room 001 (for my section)

https://act.ucsd.edu/scheduleOfClasses/scheduleOfClassesStudentResult.htm

Please double-check!

Review problems

Example: Clustering and combinatorics

- Suppose we have a dataset of 15 points, each with two features (x_1, x_2). In the dataset, there exist 3 "natural" clusters, each of which contain 5 data points.

- Recall that in the k-Means Clustering algorithm, we initialize k centroids by choosing k points at random from our dataset. Suppose $k = 3$.

Questions:

1. What's the probability that all three initial centroids are initialized in the same natural cluster?

2. What's the probability that all three initial centroids are initialized in different natural clusters?

Let S denote the whole 2-dimensional space (i.e. sample space). Let E_1, E_2, and E_3 denote the clusters (i.e. partitions):

$$E_1 \cap E_2 = \varnothing, \quad E_2 \cap E_3 = \varnothing, \quad E_3 \cap E_1 = \varnothing$$

$$E_1 \cup E_2 \cup E_3 = S.$$

Let $\mu_1, \mu_2, \mu_3 \in S$ denote the initial centroids.

For simplicity, we assume S is bounded (i.e. S is the convex hull of all $\{(x_i, y_i)\}_{i=1}^{15}$). We have:

$$P(\mu_i \in E_i) = \frac{|E_i|}{|S|}.$$

1. What's the probability that all three initial centroids are initialized in the same natural cluster?
   We further assume **uniform** distribution: $P(\mu_i \in E_i) = 1/3$.
   Because of **independence** in sampling, the final result is:

   $$P(\mu_1 \in E_1, \mu_2 \in E_2, \mu_3 \in E_3) = P(\mu_1 \in E_1)P(\mu_2 \in E_2)P(\mu_3 \in E_3) = \frac{1}{27}$$

2. What's the probability that all three initial centroids are initialized in different natural clusters?

   $$P(\mu_1 \notin E_1, \mu_2 \notin E_2, \mu_3 \notin E_3) = P(\mu_1 \notin E_1)P(\mu_2 \notin E_2)P(\mu_3 \notin E_3) = \frac{8}{27}$$

Example: basketball

Suppose we have 6 basketball players who want to organize themselves into 3 basketball teams of 2 players each. Suppose

we have three teams, “Team USA”, “Team China”, and “Team Lithuania”. How many ways can these teams be formed?

Example: basketball

Suppose we have 6 basketball players who want to organize themselves into 3 basketball teams of 2 players each. Suppose

we have three teams, “Team USA”, “Team China”, and “Team Lithuania”. How many ways can these teams be formed?

- Using permutations:

$$\frac{6!}{2!2!2!} = 90$$

- Using combinations:

$$\binom{6}{2} \cdot \binom{4}{2} \cdot \binom{2}{2} = \frac{6!}{4!2!} \cdot \frac{4!}{2!2!} \cdot 1 = \frac{6!}{2!2!2!} = 90$$

Example: basketball, again

Suppose we have 6 basketball players who want to organize themselves into 3 basketball teams of 2 players each. Now,

suppose the teams are irrelevant, and all we care about is the unique pairings themselves. How many ways can these 6 players be split into 3 teams?

Example: basketball, again

Suppose we have 6 basketball players who want to organize themselves into 3 basketball teams of 2 players each. Now,

suppose the teams are irrelevant, and all we care about is the unique pairings themselves. How many ways can these 6 players be split into 3 teams?

If the order of 3 teams matters, we have 90 ways to form the teams. If the order does not matter, we divide by 3!. Finally, we get:

$$\frac{90}{3!} = 15$$

Example: Lottery

When you buy a Powerball ticket, you select 5 different white numbers from among the numbers 1 through 59 (order of the selection does not matter), and one red number from among the numbers 1 through 35. How many different Powerball tickets can you buy?

Example: Lottery

When you buy a Powerball ticket, you select 5 different white numbers from among the numbers 1 through 59 (order of the selection does not matter), and one red number from among the numbers 1 through 35. How many different Powerball tickets can you buy?

If you check out the Powerball web site you will see that you need to select 5 distinct white numbers, so you can do this $\binom{59}{5} = 5,006,386$ ways. Then you can pick the red number $\binom{35}{1} = 35$ ways so the total number of tickets is:

$$\binom{59}{5} \cdot \binom{35}{1} = 5,006,386 \cdot 35 = 175,223,510.$$

Example: Mixed counting problems

An bag contains 15 marbles of which 10 are red and 5 are white. 4 marbles are selected from the bag.

(a) How many (different) samples (of size 4) are possible?

Example: Mixed counting problems

An bag contains 15 marbles of which 10 are red and 5 are white. 4 marbles are selected from the bag.

(a) How many (different) samples (of size 4) are possible?

$$\binom{15}{4} = 1,365$$

(b) How many samples (of size 4) consist entirely of red marbles?

Example: Mixed counting problems

An bag contains 15 marbles of which 10 are red and 5 are white. 4 marbles are selected from the bag.

(a) How many (different) samples (of size 4) are possible?

$$\binom{15}{4} = 1,365$$

(b) How many samples (of size 4) consist entirely of red marbles?

$$\binom{10}{4} = 210$$

(c) How many samples have 2 red and 2 white marbles?

Example: Mixed counting problems

An bag contains 15 marbles of which 10 are red and 5 are white. 4 marbles are selected from the bag.

(a) How many (different) samples (of size 4) are possible?

$$\binom{15}{4} = 1,365$$

(b) How many samples (of size 4) consist entirely of red marbles?

$$\binom{10}{4} = 210$$

(c) How many samples have 2 red and 2 white marbles?

$$\binom{10}{2} \cdot \binom{5}{2} = 45 \cdot 10 = 450$$

Example: Mixed counting problems

(d) How many samples (of size 4) have exactly 3 red marbles?

Example: Mixed counting problems

(d) How many samples (of size 4) have exactly 3 red marbles?

$$\binom{10}{3} \cdot \binom{5}{1} = 120 \cdot 5 = 600$$

(e) How many samples (of size 4) have at least 3 red?

Example: Mixed counting problems

(d) How many samples (of size 4) have exactly 3 red marbles?

$$\binom{10}{3} \cdot \binom{5}{1} = 120 \cdot 5 = 600$$

(e) How many samples (of size 4) have at least 3 red?

The answer is the number of samples with 3 red plus the number of samples with 4 red:

$$\binom{10}{3} \cdot \binom{5}{1} + \binom{10}{4} \cdot \binom{5}{0} = 120 \cdot 5 + 210 \cdot 1 = 810$$

Example: Mixed counting problems

(f) How many samples (of size 4) contain at least one red marble?

Example: Mixed counting problems

(f) How many samples (of size 4) contain at least one red marble?

One answer is "the number with exactly 1" + "the number with exactly 2" + "the number with exactly 3" + "the number with exactly 4":

$$\binom{10}{1} \cdot \binom{5}{3} + \binom{10}{2} \cdot \binom{5}{2} + \binom{10}{3} \cdot \binom{5}{1} + \binom{10}{4} \cdot \binom{5}{0}$$

$= 10 \cdot 10 + 45 \cdot 10 + 120 \cdot 5 + 210 \cdot 1 = 100 + 450 + 600 + 210 = 1,360$

But it is a slow computation!

Example: Mixed counting problems

(f) How many samples (of size 4) contain at least one red marble?

One answer is "the number with exactly 1" + "the number with exactly 2" + "the number with exactly 3" + "the number with exactly 4":

$$\binom{10}{1}\cdot\binom{5}{3}+\binom{10}{2}\cdot\binom{5}{2}+\binom{10}{3}\cdot\binom{5}{1}+\binom{10}{4}\cdot\binom{5}{0}$$

$$= 10\cdot 10+45\cdot 10+120\cdot 5+210\cdot 1 = 100+450+600+210 = 1,360$$

But it is a slow computation! The faster answer is the total number of samples minus the number of samples with no red marbles:

$$\binom{15}{4}-\binom{10}{0}\cdot\binom{5}{4} = 1,365-5 = 1,360.$$

Example: bitstrings

What is the probability of a randomly generated bitstring of length 5 having the same first two bits? Assume that each bit is equally likely to be a 0 or a 1.

We can start the bitstring as 00 or 11. The rest of the string does not matter. Because, there are 4 different ways to start the bitstring: $\{00, 01, 10, 11\}$. The final result is:

$$\frac{1}{2}$$

Example: bitstrings, again

What is the probability of a randomly generated bitstring of length 5 having the same first two bits, if we know that the bitstring has exactly four 0s? Assume that each bit is equally likely to be a 0 or a 1.

The first two bits must be two zeros, because there are exactly four 0s. There is also exactly one bit 1. We can only put bit 1 among the 3rd, 4th and 5th position. Thus, only 3 possibilities:

$$\{00100, 00010, 00001\}$$

The final result is:

$$\frac{3}{32}.$$

Example: Matrix inversion

Given a 2×2 matrix:

$$X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix},$$

its inverse (if exists) is given by:

$$X^{-1} = \frac{1}{\det(X)} \cdot \begin{pmatrix} x_{22} & -x_{12} \\ -x_{21} & x_{11} \end{pmatrix},$$

where $\det(X)$ is the determinant of X:

$$\det(X) = x_{11}x_{22} - x_{21}x_{12}.$$

Example: Matrix inversion

Given:

$$X = \begin{pmatrix} 5 & 2 \\ -7 & -3 \end{pmatrix},$$

what is X^{-1}?

We have the determinant:

$$\det(X) = 5 \cdot (-3) - (-7) \cdot 2 = -1,$$

and the inverse:

$$X^{-1} = \frac{1}{-1} \cdot \begin{pmatrix} -3 & -2 \\ 7 & 5 \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ -7 & -5 \end{pmatrix}.$$

Easy to verify:

$$\begin{pmatrix} 5 & 2 \\ -7 & -3 \end{pmatrix}\begin{pmatrix} 3 & 2 \\ -7 & -5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 3 & 2 \\ -7 & -5 \end{pmatrix}\begin{pmatrix} 5 & 2 \\ -7 & -3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Example: Matrix inversion

Given:

$$X = \begin{pmatrix} 5 & 2 \\ 5 & 2 \end{pmatrix},$$

what is X^{-1}?

We have the determinant:

$$\det(X) = 5 \cdot 2 - 5 \cdot 2 = 0,$$

thus there is no inverse for X.

Example: Matrix inversion

Given:

$$X = \begin{pmatrix} 5 & 5 \\ -7 & -7 \end{pmatrix},$$

what is X^{-1}?

We have the determinant:

$$\det(X) = 5 \cdot (-7) - (-7) \cdot 5 = 0,$$

thus there is no inverse for X.

Example: Linear regression by pseudo-inverse

Apply linear regression in matrix form of this data

$$D = \{(0, 0), (1, 1)\}.$$

Fundamentally, we fit a line.

We have:

$$X = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \text{ and } \vec{y} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The closed-form formula for linear regression is:

$$\vec{w} = (X^T X)^{-1} X^T \vec{y}.$$

Example: Linear regression by pseudo-inverse

First, we compute the covariance matrix:

$$X^T X = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

We have $\det(X^T X) = 2 \cdot 1 - 1 \cdot 1 = 1$, and:

$$(X^T X)^{-1} = \frac{1}{1} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

The pseudo-inverse of X is:

$$(X^T X)^{-1} X^T = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}.$$

The final result for $\vec{w}$ is:

$$\vec{w} = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix} = (X^T X)^{-1} X^T \vec{y} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Example: Sigmoid

We have the sigmoid function:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Let's derive its' derivative!

Example: Sigmoid

We have the sigmoid function:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Let's derive its' derivative!

$$\frac{d}{dx}\sigma(x) = \frac{d}{dx}\frac{1}{1 + e^{-x}} = \frac{d}{dx}(1 + e^{-x})^{-1} = -(1 + e^{-x})^{-2}\frac{d}{dx}(1 + e^{-x})$$

$$\Leftrightarrow \frac{d}{dx}\sigma(x) = (1 + e^{-x})^{-2}e^{-x} = \frac{e^{-x}}{(1 + e^{-x})^2} = \frac{1}{1 + e^{-x}} \cdot \frac{e^{-x}}{1 + e^{-x}}$$

$$\Leftrightarrow \frac{d}{dx}\sigma(x) = \sigma(x) \cdot \frac{e^{-x} + 1 - 1}{1 + e^{-x}} = \sigma(x) \cdot [1 - \sigma(x)].$$

Example: Sigmoid

We have the sigmoid function:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Show that $\sigma(x)$ is a monotonically increasing function and $0 < \sigma(x) < 1$.

Example: Sigmoid

We have the sigmoid function:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Show that $\sigma(x)$ is a monotonically increasing function and $0 < \sigma(x) < 1$.

We have:

$$\sigma'(x) = \frac{e^{-x}}{(1 + e^{-x})^2} > 0, \quad \forall x \in (-\infty, +\infty),$$

because $e^{-x} > 0$ and $(1 + e^{-x})^2 > 0$. Thus $\sigma(x)$ is a monotonically increasing function.

Example: Sigmoid

We have $1 + e^{-x} > 0$, thus $\sigma(x) > 0$. Furthermore:

$$\sigma(x) < 1 \Leftrightarrow 1 < 1 + e^{-x} \Leftrightarrow e^{-x} > 0$$

holds. Since:

$$\lim_{x \to +\infty} e^{-x} = 0, \qquad \lim_{x \to -\infty} e^{-x} = +\infty,$$

we get:

$$\lim_{x \to -\infty} \sigma(x) = 0,$$

$$\lim_{x \to +\infty} \sigma(x) = 1.$$

Example: Logistic

Given a hypothesis

$$h(\vec{x}; \vec{w}, w_0) = \sigma(\vec{x} \cdot \vec{w} + w_0).$$

In logistic regression, we have:

$$p(y = 1 | \vec{x}; \vec{w}, w_0) = \sigma(\vec{x} \cdot \vec{w} + w_0).$$

Let's derive the partial derivative of h with respect to $\vec{w}$ and w_0!

Example: Logistic

First of all, we have:

$$\frac{d}{dw_0}(\vec{x}\cdot\vec{w}+w_0)=1,$$

$$\frac{\partial}{\partial\vec{w}}(\vec{x}\cdot\vec{w}+w_0)=\vec{x}.$$

For convenience, let $\bar{h}=\vec{x}\cdot\vec{w}+w_0$. By the chain rule, we get:

$$\frac{dh}{dw_0}=\sigma'(\bar{h})\cdot\frac{d\bar{h}}{dw_0}=\sigma(\bar{h})\cdot[1-\sigma(\bar{h})],$$

$$\frac{\partial h}{\partial\vec{w}}=\sigma'(\bar{h})\cdot\frac{\partial\bar{h}}{\partial\vec{w}}=\sigma(\bar{h})\cdot[1-\sigma(\bar{h})]\cdot\vec{x}.$$