Math diagnostic

Michael Zhang and Chandra Gummaluru

CSC311, Winter 2023

Due Jan 20, 2023 5pm

This assignment will be graded for a good faith effort. The main purpose is to help you decide whether you have the math background required for the course. This is not meant to be a difficult assignment.

1/12 - added clarification on problem 6.

1. Consider a coin that lands heads with probability p. The entropy of a coin flip is defined as

$$-p\log_2 p - (1-p)\log_2(1-p)$$

   What is the value of p which maximizes the entropy? Give an 1-2 sentence intuitive explanation of why the entropy is considered a measure of uncertainty.

2. Let X be a random variable that takes the value 1 on heads and 0 on tails on a fair coin flip. Compute $\mathbb{E}[X]$ and $Var(X)$.

3. Consider square n-dimensional matrices A, B, and C.

   - Is it true that $AB = BA$?
   - Under what conditions does $AB = AC \to B = C$?

4. What does it mean for a set of vectors to form a basis for $\mathbb{R}^3$? Please specify the two necessary conditions.

5. Give an example of a matrix which has eigenvalues of 23, 20, and 1.

6. Let $x \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$. What are the gradient and Hessian of x^TAx with respect to x? You can assume A is symmetric if it simplifies your calculations.

1. Consider a coin that lands heads with probability p. The entropy of a coin flip is defined as

$$-p\log_2 p-(1-p)\log_2(1-p)$$

What is the value of p which maximizes the entropy? Give an 1-2 sentence intuitive explanation of why the entropy is considered a measure of uncertainty.

(1) let $f(p) = -p\log_2 p - (1-p)\log_2(1-p)$

Then

$$f'(p) = -\log_2 p - \frac{p}{p}\cdot \ln 2 - \left[(-1)\cdot \log_2(1-p) - \frac{1-p}{1-p}\cdot \ln 2\right]$$

$$= -\log_2 p - \ln 2 - \left[-\log_2(1-p) - \ln 2\right]$$

$$= -\log_2 p - \ln 2 + \log_2(1-p) + \ln 2$$

$$= \log_2 \frac{1-p}{p}$$

Then

$$f'(p) = 0$$

$$\Leftrightarrow \log_2 \frac{1-p}{p} = 0$$

$$\Leftrightarrow \frac{1-p}{p} = 1$$

$$\Leftrightarrow p = 0.5$$

①

Next, check the second derivative of $f(p)$ with respect to p.

$$f''(p) = \left[\log_2\left(\frac{1-p}{p}\right)\right]'$$

$$= \frac{p}{1-p}\cdot\frac{-1\cdot p-(1-p)}{p^2}\cdot \ln 2$$

$$= \frac{\ln 2}{p-1} \quad \text{when } p \neq 0$$

$\Rightarrow f''(0.5) = \ln\frac{1}{4} < 0$ ②

①, ② $\Rightarrow$ f attain the maximum value at $p = 0.5$

(2) In this case, the closer the p is to 0 or 1, the less uncertainty the outcome of flipping a coin will be, the further the p is from 0 or 1, the more such the uncertainty is. This behaves in the same way as the relationship between the entropy and p, implying that the entropy sort of measures the outcome uncertainty.

2. Let X be a random variable that takes the value 1 on heads and 0 on tails on a fair coin flip. Compute $\mathbb{E}[X]$ and $Var(X)$.

$$\mathbb{E}[X] = \mathbb{P}(X=1) \cdot 1 + \mathbb{P}(X=0) \cdot 0$$

$$= \mathbb{P}(X=1) \cdot 1$$

$$= 0.5 \cdot 1 \text{ , since the coin is fair}$$

$$= 0.5$$

$$Var[X] = \mathbb{E}[X^2] - \mathbb{E}^2[X]$$

- $\mathbb{E}[X^2] = \sum_{x \in \{0,1\}} x^2 \cdot \mathbb{P}(x)$

$$= 0^2 \cdot \mathbb{P}(x) + 1^2 \cdot \mathbb{P}(1)$$

$$= 1^2 \cdot \mathbb{P}(1)$$

$$= 1 \cdot 0.5$$

$$= 0.5$$

$$= 0.5 - 0.5^2 = 0.25$$

3. Consider square n-dimensional matrices A, B, and C.

- Is it true that $AB = BA$?
- Under what conditions does $AB = AC \rightarrow B = C$?

- NO.

Counterexample:

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

$B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$

Then

$A \cdot B = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$

$B \cdot A = \begin{bmatrix} 23 & 34 \\ 31 & 46 \end{bmatrix}$

Thus, $AB \neq BA$

- A is invertible and A^{-1} exists.

In this case, <u>assume $AB = AC$.</u>

Then

$$AB = AC$$

$$\Rightarrow A^{-1}(AB) = A^{-1}(AC)$$

$$\Rightarrow (A^{-1}A)B = (A^{-1}A)C$$

$$\Rightarrow I \cdot B = I \cdot C$$

$$\Rightarrow \underline{B = C}$$

4. What does it mean for a set of vectors to form a basis for $\mathbb{R}^3$? Please specify the two necessary conditions.

(1) Those vectors are linearly independent

(2) None of those vectors is zero vector.

5. Give an example of a matrix which has eigenvalues of 23, 20, and 1.

$$A = \begin{bmatrix} 23 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

6. Let $x \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$. What are the gradient and Hessian of $x^T A x$ with respect to x? You can assume A is symmetric if it simplifies your calculations.

Let $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$, $X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

Let $f = x^T \cdot A \cdot X$. Then

$$f(\vec{x}) = ax_1^2 + ex_2^2 + ix_3^2 + 2bx_1x_2 + 2cx_1x_3 + 2hx_2x_3$$

$$\text{Hessian} = \nabla f = \begin{bmatrix} 2ax_1 + 2bx_2 + 2cx_3 \\ 2ex_2 + 2bx_1 + 2hx_3 \\ 2ix_3 + 2cx_1 + 2hx_2 \end{bmatrix}$$

$$= 2 \cdot A \cdot \vec{X}$$