CSE 575: Statistical Machine Learning (Spring 2021)
Instructor: Nupur Thakur

Mathematical Foundations

Arizona State University

Table of contents

Basic Linear Algebra

- Given a vector **x** of m dimensions, the transpose $\mathbf{x}^t$ is -

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 & . & . & . & . & . & . & x_m \end{bmatrix} \quad \mathbf{x}^t = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ . \\ . \\ . \\ . \\ x_m \end{bmatrix}$$

Basic Linear Algebra - Determinant

- Given a 2x2 matrix **A**, the determinant **|A|** is defined as -

$$\mathrm{A} = \begin{bmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{bmatrix} = a_{00}a_{11} - a_{01}a_{10}$$

Basic Linear Algebra - Determinant

- Given a 2x2 matrix **A**, the determinant **|A|** is defined as -

$$\mathrm{A} = \begin{bmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{bmatrix} = a_{00}a_{11} - a_{01}a_{10}$$

- Given a 3x3 matrix **A**, the determinant **|A|** is defined as -

$$\mathrm{A} = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{bmatrix} = a_{00}(a_{11}a_{22} - a_{12}a_{21}) - a_{01}(a_{10}a_{22} - a_{20}a_{12}) + a_{02}(a_{10}a_{21} - a_{11}a_{20})$$

Determinant - Examples

- What is the determinant of A = $\begin{bmatrix} 1 & 7 \\ 2 & -4 \end{bmatrix}$

Determinant - Examples

- What is the determinant of A = $\begin{bmatrix} 1 & 7 \\ 2 & -4 \end{bmatrix}$

$$|A| = 1 \times -4 - 7 \times 2 = \text{-}18$$

Determinant - Examples

- What is the determinant of A = $\begin{bmatrix} 1 & 7 \\ 2 & -4 \end{bmatrix}$

$$|A| = 1 \times -4 - 7 \times 2 = -18$$

- What is the determinant of A = $\begin{bmatrix} 2 & 3 & 7 \\ -3 & 4 & 0 \\ 1 & -1 & 6 \end{bmatrix}$

Determinant - Examples

- What is the determinant of A = $\begin{bmatrix} 1 & 7 \\ 2 & -4 \end{bmatrix}$

$$|A| = 1 \times -4 - 7 \times 2 = \text{-}18$$

- What is the determinant of A = $\begin{bmatrix} 2 & 3 & 7 \\ -3 & 4 & 0 \\ 1 & -1 & 6 \end{bmatrix}$

$$|A| = 2(4 \times 6 - 0 \times -1) - 3(-3 \times 6 - 1 \times 0) + 7(-3 \times -1 - 4 \times 1) = 95$$

Basic Linear algebra - Inverse of a matrix

- For a matrix A, inverse is denoted by A^{-1}

Basic Linear algebra - Inverse of a matrix

- For a matrix A, inverse is denoted by A^{-1}

- $AA^{-1} = A^{-1}A = I$

Basic Linear algebra - Inverse of a matrix

- For a matrix A, inverse is denoted by A^{-1}
- $AA^{-1} = A^{-1}A = I$
- Why is an inverse of a matrix needed?
    - Because matrices cannot be divided!

Inverse of a matrix - Example 1

Q. Calculate the inverse of P = $\begin{bmatrix} 1 & 2 \\ 3 & -5 \end{bmatrix}$

Inverse of a matrix - Example 1

Q. Calculate the inverse of P = $\begin{bmatrix} 1 & 2 \\ 3 & -5 \end{bmatrix}$

Step 1: Calculate the determinant. Here, |P| = -11

Inverse of a matrix - Example 1

Q. Calculate the inverse of P = $\begin{bmatrix} 1 & 2 \\ 3 & -5 \end{bmatrix}$

Step 1: Calculate the determinant. Here, |P| = -11

Step 2: Swap diagonal elements and add negative sign to off-diagonal elements.

$$\begin{bmatrix} -5 & -2 \\ -3 & 1 \end{bmatrix}$$

Inverse of a matrix - Example 1

Q. Calculate the inverse of P = $\begin{bmatrix} 1 & 2 \\ 3 & -5 \end{bmatrix}$

Step 1: Calculate the determinant. Here, |P| = -11

Step 2: Swap diagonal elements and add negative sign to off-diagonal elements.

$$\begin{bmatrix} -5 & -2 \\ -3 & 1 \end{bmatrix}$$

Step 3: Divide the matrix by the determinant.

$$\begin{bmatrix} \frac{5}{11} & \frac{2}{11} \\ \frac{3}{11} & \frac{-1}{11} \end{bmatrix}$$

Inverse of a matrix - Example 2

Q. Calculate the inverse of $\begin{bmatrix} 3 & -9 \\ 2 & -6 \end{bmatrix}$

Inverse of a matrix - Example 2

Q. Calculate the inverse of $\begin{bmatrix} 3 & -9 \\ 2 & -6 \end{bmatrix}$

Solution. Here, |A| = 0. Therefore, the inverse does not exist!

Such a matrix is called **singular matrix**!

Probability

- A probability space is a triplet $(\Omega, \mathcal{B}, P)$ that is used to model a process or an experiment with random outcomes.

Probability

- A probability space is a triplet (Ω, $\mathcal{B}$, P) that is used to model a process or an experiment with random outcomes.
    - Ω - Sample space

Probability

- A probability space is a triplet (Ω, $\mathcal{B}$, P) that is used to model a process or an experiment with random outcomes.
  - Ω - Sample space
  - $\mathcal{B}$ - Collection of subsets of Ω

Probability

- A probability space is a triplet (Ω, $\mathcal{B}$, P) that is used to model a process or an experiment with random outcomes.
  - Ω - Sample space
  - $\mathcal{B}$ - Collection of subsets of Ω
  - P - probability

Conditional Probability

- Let $(\Omega, \mathscr{B}, P)$ be a probability space and let $H \epsilon \mathscr{B}$ with $P(H)>0$. For any $B \epsilon \mathscr{B}$, $P(B|H)$ is defined as-

$$P(B|H) = P(BH) / P(H)$$

and call $P(B|H)$ the conditional probability of B given H.

Total Probability Rule

- Let $(\Omega, \mathscr{B}, P)$ be a probability space, and let $\{H_j\}$ be pairwise disjoint events in $\mathscr{B}$ (i.e. $H_j H_k = \phi$, $\forall j \neq k$) and $\cup_{[j=1,\ldots,\infty]} \mathrm{H}_\mathrm{j} = \Omega$.

  Suppose $P(H_j)>0$, $\forall j$, then,

$$P(B) = \sum_{j=1,\ldots,\infty} P(H_j)P(B|H_j)$$

Bayes Theorem

- Let $(\Omega, \mathscr{B}, P)$ be a probability space, and let $\{H_j\}$ be pairwise disjoint events in $\mathscr{B}$ (i.e. $H_jH_k = \phi,\ \forall j\neq k$) and $\cup_{[j=1,\ldots,\infty]} \mathrm{H}_{\mathrm{j}} = \Omega$. and $P(H_j)>0,\ \forall j.$ We have, $\forall B \epsilon \mathscr{B}$ and $P(B)>0,$

Bayes Theorem

- Let $(\Omega, \mathscr{B}, P)$ be a probability space, and let $\{H_j\}$ be pairwise disjoint events in $\mathscr{B}$ (i.e. $H_j H_k = \phi$, $\forall j \neq k$) and $\cup_{[j=1,\ldots,\infty]} \mathrm{H}_j = \Omega$. and $P(H_j)>0, \forall j$. We have, $\forall B \epsilon \mathscr{B}$ and $P(B)>0$,

$$P(H_j|B) = \frac{P(H_j)\, P(B\,|H_j)}{\sum_{i=1,\ldots,\infty} P(H_i)P(B|H_i)}, \quad \forall j$$

Bayes Theorem

- Let $(\Omega, \mathscr{B}, P)$ be a probability space, and let $\{H_j\}$ be pairwise disjoint events in $\mathscr{B}$ (i.e. $H_j H_k = \phi,\ \forall j \neq k$) and $\cup_{[j=1,\ldots,\infty]} \mathrm{H}_j = \Omega$. and $P(H_j) > 0,\ \forall j$. We have, $\forall B \in \mathscr{B}$ and $P(B) > 0$,

$$\text{Posterior} \Rightarrow P(H_j|B) = \frac{\overset{\text{Prior}}{P(H_j)}\ \overset{\text{Likelihood}}{P(B\,|H_j)}}{\underbrace{\sum_{i=1,\ldots,\infty} P(H_i)P(B|H_i)}_{\text{Evidence}}}, \quad \forall j$$

Bayes Theorem

Consider two events A and B, then the joint probability is-

$$P(AB) = P(B|A)P(A)$$

$$\Rightarrow \quad P(AB) = P(A|B)P(B)$$

$$\Rightarrow \quad P(B|A)P(A) = P(A|B)P(B)$$

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

Bayes Theorem - Example 1

Q. A test is developed to detect a disease that 0.1% of the population have. The test is 99% effective in detecting an infected person. However, it gives a false positive result for 0.5% of cases. Find the probability that a person actually has the disease if the person tests positive?

Sol.: Let X be the event that a person has the disease & Y be the event that the test result is true.

P(X) = 0.001, P(Y|X) = 0.99, P(Y|~X) = 0.005, P(~X)=1-P(X)=0.991. We need to find P(X|Y).

Using Bayes theorem and total probability rule, we have

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y|X)P(X)+P(Y|\sim X)P(\sim X)}$$

Substituting the above values, we get P(X|Y) = 0.165

Questions?