Assignment 1: Linear Algebra; Convex Optimization; Linear Regression

Benny Chen

September 22, 2023

1 Linear Algebra and Probability

1.1 Given the following four vectors

$x_1 = [0, 0.2, 1.0, 2.2]$
$x_2 = [0.7, 0.2, 0.5, 2.0]$
$x_3 = [0, 1.0, 1.5, 2.2]$
$x_4 = [0.8, 0.1, 1.2, 2.0]$

Which point is closest to x_1 under each of the following norms?

(a) L_0

(b) L_1

(c) L_2

(d) L_∞

Answer:

(a)

x_3 is closest to x_1 under L_0 norm.
$(x_1, x_2) = |0 - 0.7|^0 + |0.2 - 0.2|^0 + |1.0 - 0.5|^0 + |2.2 - 2.0|^0 = 3$
$(x_1, x_3) = |0 - 0|^0 + |0.2 - 1.0|^0 + |1.0 - 1.5|^0 + |2.2 - 2.2|^0 = 2$
$(x_1, x_4) = |0 - 0.8|^0 + |0.2 - 0.1|^0 + |1.0 - 1.2|^0 + |2.2 - 2.0|^0 = 4$

(b)

x_3 and x_4 are closest to x_1 under L_1 norm.
$(x_1, x_2) = |0 - 0.7|^1 + |0.2 - 0.2|^1 + |1.0 - 0.5|^1 + |2.2 - 2.0|^1 = 1.4$
$(x_1, x_3) = |0 - 0|^1 + |0.2 - 1.0|^1 + |1.0 - 1.5|^1 + |2.2 - 2.2|^1 = 1.3$
$(x_1, x_4) = |0 - 0.8|^1 + |0.2 - 0.1|^1 + |1.0 - 1.2|^1 + |2.2 - 2.0|^1 = 1.3$

(c)

x_4 is closest to x_1 under L_2 norm.
$(x_1, x_2) = \sqrt{|0-0.7|^2 + |0.2-0.2|^2 + |1.0-0.5|^2 + |2.2-2.0|^2} = 0.78$
$(x_1, x_3) = \sqrt{|0-0|^2 + |0.2-1.0|^2 + |1.0-1.5|^2 + |2.2-2.2|^2} = 0.89$
$(x_1, x_4) = \sqrt{|0-0.8|^2 + |0.2-0.1|^2 + |1.0-1.2|^2 + |2.2-2.0|^2} = 0.73$

(d)

x_2 is closest to x_1 under L_∞ norm.
$(x_1, x_2) = \max\{|0-0.7|, |0.2-0.2|, |1.0-0.5|, |2.2-2.0|\} = 0.7$
$(x_1, x_3) = \max\{|0-0|, |0.2-1.0|, |1.0-1.5|, |2.2-2.2|\} = 0.8$
$(x_1, x_4) = \max\{|0-0.8|, |0.2-0.1|, |1.0-1.2|, |2.2-2.0|\} = 0.8$

1.2

(4pt) If $X \sim N(\mu, \sigma^2)$, E[X] = μ, Var[X] =σ^2, and $E[X^2] = \mu^2 + \sigma^2$. Also, recall that expectation is linear, so it obeys the following three properties:

$E[X+c] = E[X] + c$ for any constant c,
$E[X+Y] = E[X] + E[Y]$,
$E[aX] = aE[X]$ for any constant a.

We note that if X and X' are independent, then $E[XX'] = E[X]E[X']$.
Consider two points (sampled independently) from the same class follow: $X \sim N(\mu_1, \sigma^2)$ and $X' \sim N(\mu_1, \sigma^2)$.

What is the expected squared distance between them, i.e., $E[(X - X')^2]$?

Answer:

$$E[(X - X')^2] = E[X^2 - 2XX' + X'^2] \tag{1}$$

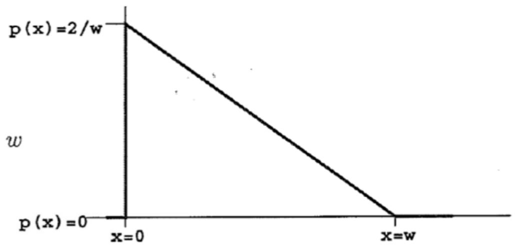
$$E[X^2 - 2XX' + X'^2] = E[X^2] - 2E[XX'] + E[X'^2] \tag{2}$$

$$\mu^2 + \sigma^2 - 2\mu^2 + \mu^2 + \sigma^2 = 2\sigma^2 \tag{3}$$

1.3

(4pt) Consider the probability density function shown in the following figure and equations.

$$
\begin{aligned}
p(x) &= \quad 0 \qquad\quad \text{if } x < 0 \\
p(x) &= \frac{2}{w} - \frac{2x}{w^2} \quad \text{if } 0 \le x \le w \\
p(x) &= \quad 0 \qquad\quad \text{if } w < x
\end{aligned}
$$



(2pt) Which *one* of the following expressions is true?

(a) $E[X] = \int_{x=-\infty}^{\infty} w(\frac{2}{w} - \frac{2x}{w^2})dx$

(b) $E[X] = \int_{x=0}^{w} x(\frac{2}{w} - \frac{2x}{w^2})dx$

(c) $E[X] = \int_{x=0}^{w} w(\frac{2}{w} - \frac{2x}{w^2})dx$

(d) $E[X] = \int_{x=-\infty}^{\infty} (\frac{2}{w} - \frac{2x}{w^2})dx$

1.3.1 Answer:

B is the correct expression
What is $p(x = 1|w = 2)$?

1.3.2 Answer:

$$p(x = 1|w = 2) = \frac{2}{2} - \frac{(2)(1)}{(2)^2} = \frac{1}{2} \tag{4}$$

1.4

Consider a feature x which is a continuous random variable with possible outcomes being all the nonnegative real numbers. The random variable follows a distribution with the following probability density function (PDF):

$$p(x|\lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \ge 0 \\ 0 & x < 0 \end{cases} \tag{5}$$

where the parameter λ of the distribution is a positive real number. Given a data set X $= \{x_1, x_2, \ldots, x_n\}$ drawn independent and identically distributed (i.i.d.) from the distribution, derive the maximum likelihood estimate (MLE) of λ based on X.

1.4.1 Answer:

$$p(X|\lambda) = \prod_{i=1}^{n} p(x_i|\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} \tag{6}$$

$$\ln p(X|\lambda) = \ln \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i \tag{7}$$

$$\frac{\partial}{\partial \lambda} \ln p(X|\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \tag{8}$$

$$\frac{n}{\lambda} = \sum_{i=1}^{n} x_i \tag{9}$$

$$\lambda = \frac{n}{\sum_{i=1}^{n} x_i} \tag{10}$$

2 Introduction to Optimization

2.1

Please justify if the following statement is correct or not (2 pt for correct T/F 3 pt each for correct explanation).

(a) In machine learning, the optimization problem we are solving to train a model is always a maximization problem.

(b) For any constant $c \in R$, $f(x) = \frac{x^2}{c-x}$ is convex on $-\infty < x < c$.

2.1.1 Answer:

(a) False. In machine learning, the optimization problem we are solving to train a model is not always a maximization problem. It can be a minimization problem as well.

(b) True. $f(x) = \frac{x^2}{c-x}$ is convex on $-\infty < x < c$.

$$f''(x) = \frac{2(c-x) - 2x}{(c-x)^2} = \frac{2c - 2x}{(c-x)^2} > 0 \tag{11}$$

2.2

Use the method of Lagrange multipliers to find the maximum values of the objective function. Please provide the maximum values and the corresponding variables. objective function:
Maximize $f(x, y) = 6xy$ subject to $\frac{x^2}{9} + \frac{y^2}{16} = 1$

2.2.1 Answer:

$$\frac{x^2}{9} + \frac{y^2}{16} = 1 => g(x,y) = \frac{x^2}{9} + \frac{y^2}{16} - 1 = 0 \tag{12}$$

$$L(x,y,\lambda) = f(x,y) + \lambda g(x,y) = 6xy + \lambda(\frac{x^2}{9} + \frac{y^2}{16} - 1) \tag{13}$$

$$\frac{\partial}{\partial x} L(x,y,\lambda) = 6y + \frac{2\lambda x}{9} = 0 \tag{14}$$

$$\frac{\partial}{\partial y} L(x,y,\lambda) = 6x + \frac{2\lambda y}{16} = 0 \tag{15}$$

$$\frac{\partial}{\partial \lambda} L(x,y,\lambda) = \frac{x^2}{9} + \frac{y^2}{16} - 1 = 0 \tag{16}$$

Solving as a system of equations would get the values:

$$(x,y) = (\frac{3\sqrt{2}}{2}, 2\sqrt{2}) \tag{17}$$

$$(x,y) = (-\frac{3\sqrt{2}}{2}, -2\sqrt{2}) \tag{18}$$

which gives us 36 as the maximum value.

3 Linear Regression

3.1

Given known $X \in R^{n \times d}$, $y \in R^{n \times 1}$, and unknown $w \in R^{d \times 1}$, $y = Xw + \epsilon$, where $\epsilon\ N(0, \sigma^2 I)$. The task is to estimate w.

(a) Please write down the loss function for the linear regression. Then derive the closed form estimation for w based on the lease square method. Note that the derive process is required and we assume that $X^T X$ is invertible, i.e, $(X^T X)^{-1}$ exists.

(b) Given X = $\begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 1 & 3 \end{bmatrix}$ and y = $\begin{bmatrix} 5 \\ 3 \\ 2 \end{bmatrix}$. Using the closed form estimation for w based on the least square method, please compute $X^T X$, $X^T y$ and the estimated w.

3.1.1 Answer:

(a)

The loss function for linear regression is:

$$L(\hat{w}) = (y - X\hat{w})^T (y - X\hat{w}) \tag{19}$$

To derive the closed form estimation for w based on the least square method, we need to take the derivative of the loss function and set it to 0.

$$\frac{\partial}{\partial \hat{w}} L(\hat{w}) = \frac{\partial}{\partial \hat{w}} (y - X\hat{w})^T (y - X\hat{w}) = 0 \tag{20}$$

$$\frac{\partial}{\partial \hat{w}} (y - X\hat{w})^T (y - X\hat{w}) = -2X^T (y - X\hat{w}) = 0 \tag{21}$$

$$X^T y = X^T X \hat{w} \tag{22}$$

$$\hat{w} = (X^T X)^{-1} X^T y \tag{23}$$

(b)

$$X^T X = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 8 \\ 8 & 14 \end{bmatrix} \tag{24}$$

$$X^T y = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 13 \\ 17 \end{bmatrix} \tag{25}$$

Using the closed form estimation for w based on the least square method, we get:

$$\hat{w} = (X^T X)^{-1} X^T y1 = \begin{bmatrix} 6 & 8 \\ 8 & 14 \end{bmatrix}^{-1} \begin{bmatrix} 13 \\ 17 \end{bmatrix} = \begin{bmatrix} \frac{23}{10} \\ \frac{-1}{10} \end{bmatrix} \tag{26}$$