

Homework 0

Helen Ngo

October 11, 2016

User names Github: helenngo

Kaggle: idkngo (Display name: HelenNgo)

Problem 1 For the function $g(x) = -3x^2 + 24x - 30$, find the value for x that maximizes $g(x)$.

Solution: All relative extrema can be found at the critical values of the function g where the derivative $g'(x) = 0$. Furthermore, one can determine if a critical value is a maxima using the second derivative test[1]:

”Suppose $f(x)$ is a function of x that is twice differentiable at a stationary point x_0 .

1. If $f''(x_0) > 0$, then f has a local minimum at x_0 .
2. If $f''(x_0) < 0$, then f has a local maximum at x_0 .”

The derivative of $g(x)$ is $g'(x) = -6x + 24$, which can be rewritten as

$$g'(x) = -6(x - 4).$$

Therefore, $x = 4$ is the only critical value. Then the second derivative is $g''(x) = -6$, $g''(4) = -6 < 0$ and by the second derivative test we can determine that at $x = 4$, the function $g(x)$ is at a maxima. Since $x = 4$ is a maxima and the only critical value of the function it is the global maxima, the value $x = 4$ maximizes $g(x)$.

Problem 2 Consider the following function:

$$f(x) = 3x_0^3 - 2x_0x_1^2 + 4x_1 - 8,$$

what are the partial derivatives of $f(x)$ with respect to x_0 and x_1 .

Solution: In finding the partial derivative with respect of x_i , all other variables are treated as constants. Thus,

$$f_{x_0}(x) = 9x_0^2 - 2x_1^2$$

and

$$f_{x_1}(x) = 4x_0x_1 + 4.$$

Problem 3 Consider the matrix $A = \begin{bmatrix} 1 & 4 & -3 \\ 2 & -1 & 3 \end{bmatrix}$ and $B = M = \begin{bmatrix} -2 & 0 & 5 \\ 0 & -1 & 4 \end{bmatrix}$, then answer the following and verify your answers in Python.

- (a) Can you multiply the two matrices? Elaborate on your answer.
- (b) Multiply A^T and B and give its rank.
- (c) Let $C = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ be a new matrix; what is the result of $AB^T + C^{-1}$?

Solution:

- (a) In matrix multiplication, the row elements in the first matrix is combined with the corresponding column elements in the second matrix, such that column dimension of the first matrix must match the row dimension of the second matrix. Both matrices A and B have the same dimensions, but are not square matrices. Thus the matrices do not have the correct dimensions to be multiplied together.
- (b) The transpose of matrix A ,

$$A^T = \begin{bmatrix} 1 & 2 \\ 4 & -1 \\ -3 & 3 \end{bmatrix}.$$

Then

$$A^TB = \begin{bmatrix} 1 & 2 \\ 4 & -1 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} -2 & 0 & 5 \\ 0 & -1 & 4 \end{bmatrix}.$$

Using the matrix multiplication, the first element of the product matrix is $1(-2) + 2(0) = -2$.
Solving the rest of the matrix in the same way we get the 3×3 matrix

$$A^T B = \begin{bmatrix} -2 & -2 & 13 \\ -8 & 1 & 16 \\ 6 & -3 & -3 \end{bmatrix}.$$

The rank of a matrix is the number of linearly independent rows (or columns) in the matrix.
Using the "rref" function on a calculator, the reduced echelon form matrix $A^T B$ could be found as

$$\begin{bmatrix} 1 & 0 & -2.5 \\ 0 & 1 & -4 \\ 0 & 0 & 0 \end{bmatrix}.$$

It can be seen from the reduced echelon form that there are two linearly independent rows, thus the rank of $A^T B$ is 2.

(c) In the augmented matrix method, we use Gauss-Jordan elimination to transform

$$\left[\begin{array}{c|c} A & I \end{array} \right] \rightarrow \left[\begin{array}{c|c} I & A^{-1} \end{array} \right]$$

where I is the identity matrix, A is the original matrix and A^{-1} is the inverse matrix. Then

$$\left[\begin{array}{c|c} C & I \end{array} \right] = \left[\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \end{array} \right] \rightarrow \left[\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & \frac{1}{2} \end{array} \right] = \left[\begin{array}{c|c} I & C^{-1} \end{array} \right].$$

Therefore

$$\begin{aligned}
 AB^T + C^{-1} &= \begin{bmatrix} 1 & 4 & -3 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} -2 & 0 \\ 0 & -1 \\ 5 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \\
 &= \begin{bmatrix} -17 & -16 \\ 11 & 13 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \\
 &= \begin{bmatrix} -16 & -16 \\ 11 & 13\frac{1}{2} \end{bmatrix}
 \end{aligned}$$

(d) In Python, this problem could be easily verified:

```

>>> import numpy as np
>>> A = [[1,4,-3],[2,-1,3]]
>>> B = [[-2,0,5],[0,-1,4]]
>>> np.dot(A, B)
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
ValueError: shapes (2,3) and (2,3) not aligned: 3 (dim 1) != 2 (dim 0)
>>> ans = np.dot(np.transpose(A),B)
>>> ans
array([[ -2,  -2, 13],
       [-8,   1, 16],
       [ 6,  -3, -3]])
>>> from numpy.linalg import matrix_rank
>>> matrix_rank(ans)
2
>>> C = [[1,0],[0,2]]
>>> from numpy.linalg import inv
>>> D = np.dot(A,np.transpose(B))+inv(C)

```

```
>>> D
array([[ -16.  ,  -16.  ],
       [ 11.  ,   13.5]])
```

Problem 4 Give the mathematical definitions of the simple Gaussian, multivariate Gaussian, Bernoulli, binomial and exponential distributions.

Solution: A Gaussian or normal distribution[2] "in a variate X with mean μ and variance σ^2 is a statistic distribution with probability density function

$$P(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

on the domain $x \in (-\infty, \infty)$." The distribution is associated with the Central Limit Theorem.

The probability density function of the d-dimensional multivariate Gaussian[1] or normal distribution is given by

$$y = f(x, \mu, \Sigma) = \frac{1}{\sqrt{|\Sigma|(2\pi)^d}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}$$

where x and μ are 1-by-d vectors and Σ is a d-by-d symmetric positive definite matrix.

A Bernoulli trial is a random experiment with two outcomes. The probability density function, mean, variance and moment generating function of a Bernoulli Distribution[3] for a random variable X is

$$P(X = x|\pi) = \pi^x(1 - \pi)^{1-x}, x = 0, 1$$

$$E[X] = \pi$$

$$Var[X] = \pi(1 - \pi)$$

$$M_x(t) = \pi e^t + \varrho.$$

An experiment with a fixed number of n independent Bernoulli trials with the random variable X , the number of observed successes during the n trials is called a Binomial experiment. The corresponding

distribution, the binomial Distribution[4] has the following pdf, mean, variance and mgf:

$$P(X = x|n, \pi) = \binom{n}{x} \pi^x (1 - \pi)^{n-x}, x = 0, 1, 2, \dots, n.$$

$$E[X] = n\pi$$

$$Var[X] = n\pi(1 - \pi)$$

$$M_x(t) = (\pi e^t + \varrho)^n.$$

The exponential distribution[4] is "characterized by a lack of memory property and is often used to model lifetimes of electronic components as well as waiting times for Poisson processes." The pdf, mean, variance and mgf of an exponential distribution is

$$f(x|\lambda) =$$

$$E[X] = \frac{1}{\lambda}$$

$$Var[X] = \frac{1}{\lambda^2}$$

$$M_x(t) = (1 - \lambda^{-1}t)^{-1} \text{ for } t < \lambda.$$

In the definitions above, please note that π is the probability of success and λ is the mean.

Problem 5 What is the relationship between the Bernoulli and binomial distributions?

Solution: The Bernoulli distribution shows the distribution of the outcomes of Bernoulli trial, a "random experiment with only two possible outcomes".[4] The binomial distribution shows the distribution of a sequence of a fixed number of independent Bernoulli trials, and the probability of success at each trial is the same. In other words, the Bernoulli distribution considers one Bernoulli trial, and binomial distribution considers n Bernoulli trials.

Problem 6 Suppose that random variable $X \sim N(2, 3)$. What is its expected value?

Solution: The expected value of a normal distribution is its mean, $E[X] = \mu$, and a normal distribution is denoted as $X \sim N(\mu, \sigma)$. Therefore the expected value of random variable $X \sim N(2, 3)$ is 2.

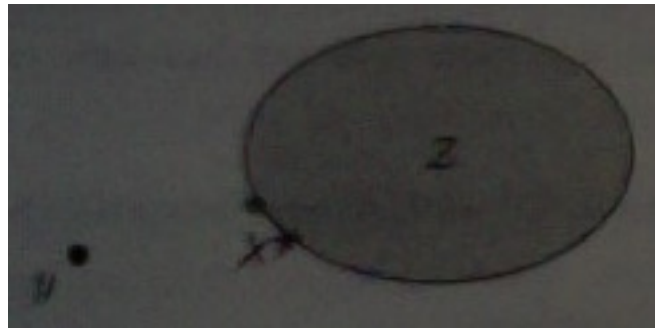
Problem 7 An euclidean projection of a d -dimensional point $y \in \mathbb{R}^d$ to a set Z is given by the following optimization problem:

$$x^* = \arg \min \|x - y\|_2^2, \text{ subject to } : x \in Z$$

where Z is the feasible set, $\|\cdot\|_2$ is the l_2 -norm (euclidian) of a vector, and $x^* \in \mathbb{R}^d$ is the projected vector.

- (a) What is x^* if $y = 1.1$ and $Z = \mathbb{N}$, where \mathbb{N} is the set of natural numbers?
- (b) Locate x^* in a picture.

Solution: The argument minimizes the distance between x and y . Since $y = 1.1$ and $x \in Z = \mathbb{N}$, $x^* = 1$. This can be determined by observation.



Problem 8 Suppose that random variable Y has distribution

$$p(Y = y) = \begin{cases} e^{-y} & \text{if } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Solution:

- (a)

$$\begin{aligned} \int_{y=-\infty}^{\infty} p(Y = y) dy &= \lim_{a \rightarrow -\infty} \int_a^0 0 dy + \lim_{b \rightarrow \infty} \int_0^b e^{-y} dy \\ &= 0 + \lim_{b \rightarrow \infty} \int_0^b e^{-y} dy \\ &= \lim_{b \rightarrow \infty} -e^{-b} + e^0 \\ &= 0 + 1 = 1 \end{aligned}$$

The sum of all probabilities for the random variable adds up to 1.

(b)

$$\begin{aligned}\mu_Y &= E[Y] \\ &= \int_{y=-\infty}^{\infty} p(Y=y)y \, dy \\ &= \lim_{a \rightarrow -\infty} \int_a^0 0y \, dy + \lim_{b \rightarrow \infty} \int_0^b e^{-y}y \, dy \\ &= 0 + \lim_{b \rightarrow \infty} -e^{-y}(y+1) \Big|_0^b \\ &= \lim_{b \rightarrow \infty} -e^{-b}(y+1) + e^0(0+1) \\ &= 1\end{aligned}$$

The expected value of Y , $E[Y] = 1$.

(c) Note that the result from part (b) was used.

$$\begin{aligned}\sigma^2 &= Var[Y] \\ &= \int_{y=-\infty}^{\infty} p(Y=y)(y-\mu_Y)^2 \, dy \\ &= \lim_{a \rightarrow -\infty} \int_a^0 0 \, dy + \lim_{b \rightarrow \infty} \int_0^b e^{-y}(y-1)^2 \, dy \\ &= 0 + \lim_{b \rightarrow \infty} \int_0^b e^{-y}(y-1)^2 \, dy \\ &= \lim_{b \rightarrow \infty} \int_0^b e^{-y}(y^2 - 2y + 1) \, dy \\ &= \lim_{b \rightarrow \infty} \int_0^b (y^2 e^{-y} - 2y e^{-y} + e^{-y}) \, dy \\ &= \lim_{b \rightarrow \infty} -e^{-y}(y^2 + 2y + 2) - 2e^{-y}(y+1) + e^{-y} \Big|_0^b \\ &= \lim_{b \rightarrow \infty} -e^{-y}(y^2 + 1) \Big|_0^b \\ &= \lim_{b \rightarrow \infty} -e^{-b}(y^2 + 1) + e^0(0^2 + 1) \\ &= 0 + 1(1) = 1\end{aligned}$$

The variance of Y , $Var[Y] = 1$.

- (d) The expected value of Y given that $Y \geq 10$ can be found with the integral in part (b) and evaluating the integral from 10 to ∞ .

$$\begin{aligned}
 \int_1^b 0^b e^{-y} y \, dy &= \lim_{b \rightarrow \infty} \int_1^b 0^b e^{-y} y \, dy \\
 &= \lim_{b \rightarrow \infty} -e^{-y}(y+1) \Big|_1^b 0^b \\
 &= \lim_{b \rightarrow \infty} -e^{-b}(b+1) + e^{-10}(10+1) \\
 &= 0 + \frac{1}{e^{10}}(11) \\
 &= \frac{11}{e^{10}} \\
 &\approx 0.00049940.
 \end{aligned}$$

References

- [1] Weisstein, Eric W. "Second Derivative Test." From MathWorld—A Wolfram Web Resource.
- [2] Weisstein, Eric W. "Normal Distribution." From MathWorld—A Wolfram Web Resource.
- [3] "Multivariate Normal Distribution." MathWorks. The MathWorks, Inc., n.d. Web. 04 Sept. 2016.
- [4] Ugarte, Mara Dolores., Ana F. Militino, and Alan T. Arnholt. Probability and Statistics with R. 2nd ed. Boca Raton, FL: CRC, 2016. Print.