

STAT 542: Homework 1

Spring 2021, by Lucie Lu (lul3)

Due: Tuesday 11:59 PM CT, Feb 2nd

Contents

About Homework 1	1
Basic calculus	1
Linear algebra	2
Statistics	2
Programming	4

About Homework 1

You will receive 100% score for HW1 as long as you submit a finished version before the deadline. The goal of HW1 is to check your prerequisite knowledge. All of this knowledge should be already covered in courses such as Stat 410, Stat 425 and other basic mathematical courses. These concepts and skills will be used extensively in our course.

- If you are not familiar with some of these concepts and skills, you are strongly encouraged to review related materials as an early preparation for this course.
- If you cannot derive the answers (closed-book) on your own even after reviewing them, then STAT432 (Basics of Statistical Learning) might be a better option, which covers similar topics but is less mathematical-oriented.

In addition, this HW also check your basic programming knowledge, such as writing a function, random seed, and also Latex. Please note that you are required to type-in all formulas in Latex form in all future homework. Failing to do so will leads to some penalty.

Basic calculus

1. Calculate the derivative of $f(x)$

- (a) $f(x) = e^x$
- (b) $f(x) = \log(1 + x)$
- (c) $f(x) = \log(1 + e^x)$

Answer: (a) $\frac{df(x)}{dx} = e^x$ (b) $\frac{df(x)}{dx} = \frac{1}{1+x}$ (c) $\frac{df(x)}{dx} = \frac{e^x}{1+e^x}$

2. Taylor expansion. Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be a twice differentiable function. Please write down the first three terms of its Taylor expansion at point $x = 1$.

Answer: The Taylor Series show that

$$f(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \frac{f'''(a)}{3!}(x - a)^3 + \dots$$

At point $x = 1$, the first three terms of the Taylor expansion is

$$f(x) = f(1) + f'(1)(x - 1) + \frac{f''(1)}{2!}(x - 1)^2$$

3. For the infinite sum $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$, where α is a positive real number, give the exact range of α such that the series converges.

Answer: A series converge only when the partial sums $\sum_{n=1}^{\infty} \frac{1}{n^\alpha} \leq M$ are bounded.

Note that $\frac{1}{2^\alpha} + \frac{1}{3^\alpha} < \frac{2}{2^\alpha}$, $\frac{1}{4^\alpha} + \frac{1}{5^\alpha} + \frac{1}{6^\alpha} + \frac{1}{7^\alpha} < \frac{4}{4^\alpha}$, and so on, then $\sum_{n=1}^{2^N-1} \frac{1}{n^\alpha} < 1 + \frac{1}{2^{\alpha-1}} + \frac{1}{4^{\alpha-1}} + \frac{1}{8^{\alpha-1}} + \dots + \frac{1}{2^{(n-1)(\alpha-1)}} < \frac{1}{1-2^{1-\alpha}}$

Then we need $\lim_{N \rightarrow \infty} \frac{1}{1-2^{1-\alpha}} = 0$, so that for $1 < \alpha < \infty$, the series converges.

Linear algebra

1. What is the eigendecomposition of a real symmetric matrix $A_{n \times n}$? Write down one form of that decomposition and explain each term in your formula. Based on these terms, derive $A^{-1/2}$.

Answer:

The symmetric matrix $A_{n \times n}$ can be factored as

$$A = Q\Lambda Q^T$$

where Q is orthogonal ($Q^{-1} = Q^T$) and Λ is diagonal (eigenvalues are in the diagonal). The decomposed matrix with eigenvectors are now orthogonal matrix.

$$A^{-1/2} = Q\Lambda^{-1/2}Q^T$$

2. What is a symmetric positive definite matrix $A_{n \times n}$? Give one of equivalent definitions and explain your notation.

Answer:

Check the quadratic form is positive.

Quadratic form is defined as follows: $X \in R^{n \times 1}$, $A \in R^{n \times n}$, quadratic form is $X^T A X$, which is a scalar value. If this is positive, then $A_{n \times n}$ is positive definite. In terms of geometric interpretation, positive definite is a bowl-shaped surface.

3. True/False. If you claim a statement is false, explain why. For two real matrices $A_{m \times n}$ and $B_{n \times m}$
- (a) $\text{Rank}(A) = \max\{m, n\}$
 - (b) If $m = n$, then $\text{trace}(A) = \sum_{i=1}^n A_{ii}$
 - (c) If A is a symmetric matrix, then all eigenvalues of A are real
 - (d) If A is a symmetric matrix, λ_1 and λ_2 are two of its eigen-values (not necessarily different) and v_1, v_2 are the corresponding eigen-vectors, then $v_1^T v_2 = 0$.
 - (e) $\text{trace}(ABAB) = \text{trace}(AABB)$

Answer:

- (a) F. $\text{Rank}(A) \leq \min\{m, n\}$
- (b) T
- (c) T
- (d) T
- (e) F. $\text{trace}(ABAB) = \text{trace}(BABA)$

Statistics

1. X_1, X_2, \dots, X_n are i.i.d. $\mathcal{N}(\mu, \sigma^2)$ random variables, where $\mu \in \mathbb{R}$ and $\sigma > 0$ is finite. Let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.
- (a) What is an unbiased estimator? Is \bar{X}_n an unbiased estimator of μ ?

- (b) What is $E[(\bar{X}_n)^2]$ in terms of n, μ, σ ?
- (c) Give an unbiased estimator of σ^2 .
- (d) What is a consistent estimator? Is \bar{X}_n a consistent estimator of μ ?

Answer:

- (a) An estimator is unbiased when it equals to the parameter. In other words, the mean of the statistic $u(X_1, X_2, \dots, X_n)$ is equal to the population's parameter μ : $E[u(X_1, X_2, \dots, X_n)] = \mu$.

\bar{X}_n is an unbiased estimator of μ .

When X_i are normally distributed, $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Then

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} n\mu = \mu$$

because \bar{X}_n is the mean of the sampling distribution X , and μ is the population's mean.

- (b) $E[(\bar{X}_n)^2]$ is derived from the following:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i^2 - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2\bar{X}x_i + \bar{X}^2) = \frac{1}{n} \sum_{i=1}^n x_i^2 - 2\bar{X} \frac{1}{n} \sum_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - 2\bar{X}^2 + \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{X}^2$$

- (c) An unbiased estimator of σ^2 :

From (b),

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{X}^2$$

$$E(\sigma^2) = E\left[\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{X}^2\right] = \frac{1}{n} E\left(\sum_{i=1}^n x_i^2\right) - E(\bar{X}^2)$$

From the alternative formulas for variance, $Var(X) = \sigma^2 = E(X^2) - \mu^2$, and $Var(\bar{X}) = \frac{\sigma^2}{n} = E(\bar{X}^2) - \mu^2$

Then

$$E(\sigma^2) = E\left[\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{X}^2\right] = \frac{1}{n} E\left(\sum_{i=1}^n x_i^2\right) - E(\bar{X}^2) = \frac{1}{n} (n\sigma^2 + n\mu^2) - \frac{\sigma^2}{n} - \mu^2 = \frac{(n-1)\sigma^2}{n}$$

So $\frac{(n-1)\sigma^2}{n}$ is a biased estimator of σ^2 .

- (d) A consistent estimator has the smaller variance while sample size goes large. The more data we collect, a consistent estimator is closer to the real population parameter. Yes, \bar{X}_n is a consistent estimator of μ .

2. Suppose $X_{p \times 1}$ is a vector of covariates, $\beta_{p \times 1}$ is a vector of unknown parameters, ϵ is the unobserved random noise and we assume the linear model relationship $y = X^T \beta + \epsilon$. Suppose we have n i.i.d. samples from this linear model, and the observed data can be written using the matrix form: $\mathbf{y}_{n \times 1} = \mathbf{X}_{n \times p} \beta_{p \times 1} + \epsilon_{n \times 1}$.

- (a) If we want estimate the unknown β using a least square method, what is the objective function $L(\beta)$ to obtain $\hat{\beta}$?
- (b) What is the solution of $\hat{\beta}$? Represent the solution using the observed data \mathbf{y} and $\mathbf{X}_{n \times p}$. Note that you may assume that $\mathbf{X}^T \mathbf{X}$ is invertible.

Answer: (a) Objective function $L(\beta) = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y - \beta_0 - \beta_1 X_i)^2$. It is also called sum of squared error (SSE). (b)

$$\sum_{i=1}^n \epsilon_i^2 = \epsilon_i^T \epsilon_i = (Y - X\beta)^T (Y - X\beta) = (Y^T - \beta^T X^T)(Y - X\beta) = Y^T Y - 2Y^T X\beta + \beta^T X^T X\beta$$

$$\frac{\partial \epsilon_i^T \epsilon_i}{\partial \beta} = -2Y^T X + 2\beta^T X^T X$$

Solve $-2Y^T X + 2\beta^T X^T X = 0$, then $\hat{\beta}^T = Y^T X (X^T X)^{-1}$, so $\hat{\beta} = (X^T X)^{-1} X^T Y$

Programming

1. Use the following code to generate a set of observations \mathbf{y} and $\mathbf{X}_{n \times p}$. Following the previously established formula, Write your own code, instead of using existing functions such as `lm()`, to solve for the least square estimator $\hat{\beta}$. If you are asked to add an intercept term β_0 into your estimation (even the true $\beta_0 = 0$ in our data generator), what should you do?

```
set.seed(1)
n = 100; p = 5
X = matrix(rnorm(n * p), n, p)
y = X %*% c(1, 0, 0, 1, -1) + rnorm(n)
```

Answer:

```
#Solve for beta
solve((t(X)%*%X))%*%t(X)%*%y
```

```
##           [,1]
## [1,]  0.88369785
## [2,] -0.01725878
## [3,] -0.02658416
## [4,]  0.90233834
## [5,] -1.05639331
```

```
#add an intercept
intercept <- matrix(1, n, 1)
X_1 = cbind(intercept, X)
solve((t(X_1)%*%X_1))%*%t(X_1)%*%y
```

```
##           [,1]
## [1,] -0.02953446
## [2,]  0.88795776
## [3,] -0.01811757
## [4,] -0.02601449
## [5,]  0.90398547
## [6,] -1.05766232
```

```
#To check my work
lm(y~X)
```

```
##
## Call:
## lm(formula = y ~ X)
##
## Coefficients:
```

## (Intercept)	X1	X2	X3	X4	X5
## -0.02953	0.88796	-0.01812	-0.02601	0.90399	-1.05766

2. Perform a simulation study to check the consistency of the sample mean estimator \bar{X}_n . Please save your random seed so that the results can be replicated by others.

- Generate a set of $n = 20$ i.i.d. observations from uniform $(0, 1)$ distribution and calculate the sample mean \bar{X}_n
- Repeat step (a) 1000 times to collect 1000 such sample means and plot them using a histogram.
- How many of such sample means (out of 1000) are at least 0.1 away from true mean parameter, which is 0.5 for uniform $(0, 1)$?
- Repeat steps (a) to (c) with $n = 100$ and $n = 500$. What conclusion can you make?

Answer: (a)

```
set.seed(31498)
X_a <- runif(20)
mean_X_a <- mean(X_a)
mean_X_a
```

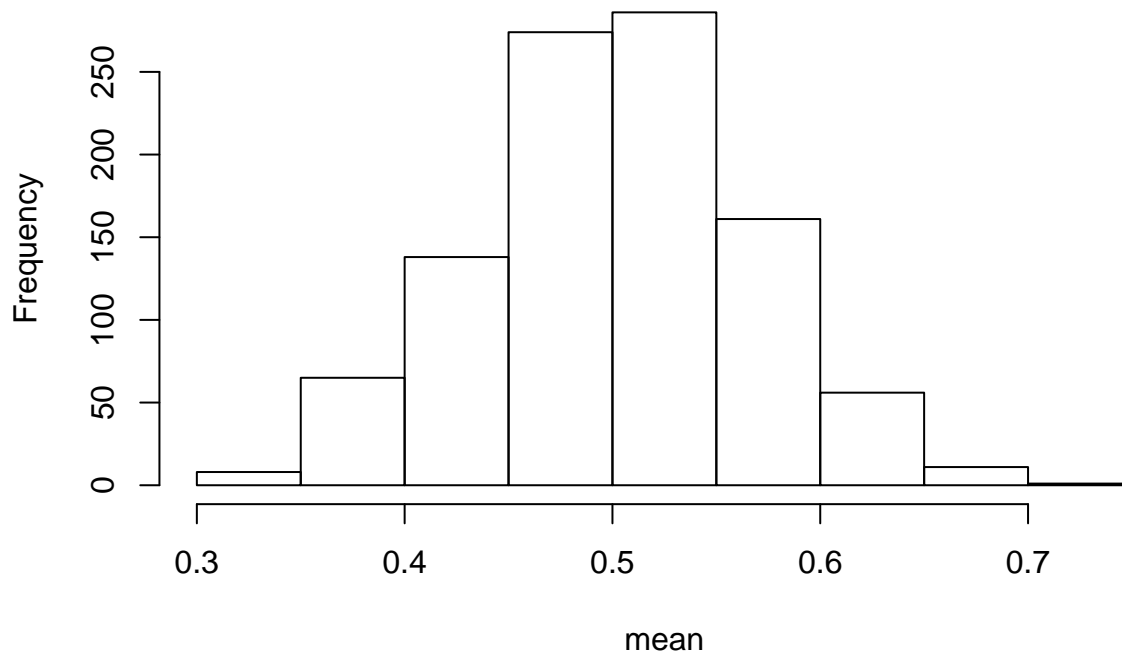
```
## [1] 0.4577221
```

$\bar{X}_n = 0.4577221$

(b)

```
set.seed(31498)
X_b <- replicate(1000, {
  X_a <- runif(20)
  mean(X_a)
})
hist(X_b, main = "Mean of random sample n=20 from uniform distribution", xlab="mean")
```

Mean of random sample n=20 from uniform distribution



(c)

```
away_20 <- summary(X_b < 0.4 | X_b > 0.6)[3]
away_20
```

```
## TRUE
## 141
```

141 out of 1000 such sample means are at least 0.1 away from the true mean.

(d)

```
set.seed(230943)
#increase n=100
X_d_100 <- replicate(1000, {
  X_d_100 <- runif(100)
  mean(X_d_100)
})
summary(X_d_100 < 0.4 | X_d_100 > 0.6)
```

```
## Mode FALSE
## logical 1000
```

```
#increase n=500
X_d_500 <- replicate(1000, {
  X_d_500 <- runif(500)
  mean(X_d_500)
})
summary(X_d_500 < 0.4 | X_d_500 > 0.6)
```

```
##      Mode    FALSE
## logical      1000
```

In both cases, when n are increased to 100 and 500, none of the 1000 sample means are at least 0.1 away from the true mean. This sample mean is a consistent estimator.