1.

a)

Expectation of slope variable in linear regression model:

Normal equation:

The linear regression model is represented as

Expectation of = E[ = E[] = E[]

= E[]= E[]=E[] = E[]+E[]

E[]=0 as follows normal distribution with mean as 0.

E[ = E[]= as all the terms are constant.

Using matrix multiplication rile ABC=CAB,

**E[**

b)

Var[]= var[] = var[]\*var[Y] = (

Var[Y]= var[] = as follows a normal distribution with variance

**Var[]=(XX’**

c)

Yes, follows Gaussian distribution

By the normal equation and follows a normal distribution with variance and mean 0.

As is linearly dependent on , also follows Gaussian distribution with mean and variance (**XX’**

d)

Covariance matrix of Y is a diagonal matrix and is represented by 𝚺

Posterior Probabilities are given by Gaussian distribution as below:

=N(ѲTx, σi2)

Likelihood function is given by Πp()=

Log likelihood function is given by

-∑i=1:N  - )2/ 2 σi2 + constant

In matrix format the above equation can be represented as

h()= (

Differentiating the log likelihood function and equating it to 0,

This implies,

2.

Ridge Regression:

Given that likelihood y ~ N(Xθ,) and prior θ~ N(0,)

Posterior distribution

The term can be considered as a normalisation constant and posterior distribution can be written as:

We know that

And

So

Where K represent all other constant terms present in the equation

Differentiating the log likelihood function w.r.t to and equating it to zero,

This implies

The equation above is similar to ridge regression expression. So the expected value of ridge regression parameter can be set as

When p(x|) is Gaussian and p( is Gaussian, then p() is Gaussian as product of two Gaussian signals is also Gaussian. So by that logic given and are Gaussian, posterior distribution will also be a Gaussian signal. Let be the mean and variance of the Gaussian posterior distribution.

\*

By substituting Gaussian equation and equating the exponentials on both the sides of equation we get,

Given the above equation mean can be estimated as:

So, the mean of the posterior distribution under a Gaussian prior

3.

3.1

Bayes rule of classification with loss functions is to minimize the risk function caused by the loss.

Given L(a,b)=p and L(b,a)=q

Let’s assume that ‘a’ corresponds to all the events or X whose Y = 1 and ‘b’ corresponds to all the events whose Y=-1

Risk function is given by

Where and correspond to taking y=1 as action and y=-1 as action respectively and correspond to the y=1 and y=-1 classes

To decide that the current ‘x’ belong to ,

If the above condition is satisfied, then x belongs to class else it belongs to .

Here

So the discriminant function can be written as

The classifier can also be expressed as

The classifier can be defined as

if g(x)<0 then x belong to y=1 class else to y=-1 class

classifier function f(X)= sign(g(x))

3.2

a)

Given that class conditional distribution is Gaussian i.e.,

and

And

Using the classifier function g(x) derived in 3.1, and plugging normal equations for class conditional distributions,

If g(x)<0, x belongs to y=1 class, else if g(x)>0 x belongs to y=-1 class.

Bayes classifier is

sign(

Decision boundary is defined as all x for g(x)=0. That is

=0

Where K is a constant and K= 2\*exp(-)

If written in discrete x terms, the above equation can be represented as

Where and

In matrix form the above equation can be represented as

The equation represents a shapes defined by quadratic sets. So the decision boundary can usually take shapes of hyperbolas or parabolas or ellipse.

b)

When both the classes have same covariance matrix, the equation can be represented as

Or = K

This implies

The above equation represents a straight line. So when covariance matrices are same decision boundary takes the shape of straight line.

c)

When the covariance matrix is an identity matrix, the equation representing decision boundary line can be represented as

Or in matrix form it can be represented as

The above equation represents a straight line. So the decision boundary is a straight line.

4.

a).

P(y=1|X=x) =

P(y=0|X=x) is given by P(y=0|X=x)=1-P(y=1|X=x) = 1- =

Log odds of success = ln(

= ln()

= ln()

**=**

The above log odd of success is a linear function of x.

b).

Logistic loss is given by

A function f(x) can be proved as a convex function if

The above expression is always greater than equal to zero as are always greater than or equal to zero. So the logistic loss function is convex.

5.

a) Root mean square error is expressed as

**This implies**

**This implies**

b) Root mean square error is expressed as

**This implies - λ\* ]**

**This implies - ]**

**Report:**

My RMSE for various lowrank

Low rank-1 : Train RMSE: 0.9264

Test RMSE:0.9547

Low rank-3 : Train RMSE: 0.8847

Test RMSE:0.9519

Low rank-5 : Train RMSE: 0.8231

Test RMSE:0.9366

With low ranks, train RMSE is high comparatively. And the computational time for lesser lowrank is less. I varied μ,λ and chose those values for which RMSE is less. Initially when my learning rate μ was very high, the model is not getting converged and the error raised upto infinity in few iterations. So I started reducing the μ value which gives lesser RMSE performance. After fixing μ, I varied λ and fixed that λ value which results in least RMSE.