**Ans1:** Here, it is given that is the input data and

Also, and is the measurement of for the -th training sample.

Now,

The cost function of the Ridge Regression is

[All the matrices are symmetric, so ]

[A and B are symmetric]

To get the closed form equation of Ridge regression differentiate and minimize the cost function,

Multiplying both sides with ,

**Hence Proved**

Here X is matrix of input, y is the measurement of for the -th training sample and I is the identity matrix.

**Ans2:**

1. To learn this SoftMax Regression model, we need to estimate parameter and the parameters are

**Diagram:**

... 1 🡪 Input features

┌───┐ ┌───┐ ┌───┐ ┌───┐

│ │ │ │ │ │ │ │

│ │ │ │ │ │ │ │

│ │ │ │ │ │ │ │

│ │ │ │ │ │ │ │

└───┘ └───┘ └───┘ └───┘

,1 ,1 ,1 ,1

Softmax

Posterior Probability

probability of probability of

class 1 () class K ()

Where

For minimizing the cost function,

We know if instance belongs to k, otherwise,

SoftMax regression gradient for cross-entropy cost function:

**Hence Proved.**