Linear models

Q1 :

Q2 :

Q3:

Now define

Also for x, when this is used in the function it gives:

Deriving this gives

Q4:

Q5:

Q6:

Nonlinear functions

Q7:

ReLu :

Sigmoid:

Softmax:

Q8:

ReLu: for all x large than zero the derivative is 1.

Sigmoid: if x gets large the derivative goes to zero

Softmax: This function is not dependent on the values of x but only on the ratios between the elements of x. Thus it will give a good gradient regardless of x values.

Shallow nonlinear models

Q9: The space of XOR cannot be divided by a line. Even if it is mapped linearly to another space it will still be linear and thus cannot be divided. When it is mapped non-linearly to another space, the points can be divided by a line.

Q10:

Q11:

Binary classification with logistic regression

Q12: A softmax function gives the probability of the input belonging to each class. Then the input will be classified as the class whit the highest probability. That is why a softmax function is useful for categorical classification.

Q13:

Q14:

Q15:

Q16:

Classification with a shallow nonlinear model

Q17: