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 approaches 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 outputs of the XOR cannot be classified effectively using a linear decision boundary. Even if multiple linear transformations are combined the resulting decision boundary will remain linear. However, by applying a non-linear transform the inputs are in essence mapped to a non-linear coordinate system. By picking the right transform we can make it easier to separate the input points using a linear decision boundary.

Q10:

Q11:

Binary classification with logistic regression

Q12: A softmax function gives the probability of the input belonging to each class. Consequently, the input can be classified by picking the class with 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: