CS551: Introduction to Deep Learning

Mid Semester, Spring 2017

IIT Patna

Attempt all questions. Do not write anything on the question paper.

Time: 2 Hrs | Full marks: 30

1. Prove that softmax is invariant to constant offsets in the input, that is, for any input vector $\boldsymbol{x}$ and any constant $\boldsymbol{c}$, softmax($\boldsymbol{x}$)=softmax($\boldsymbol{x}+\boldsymbol{c}$), where $\boldsymbol{x}+\boldsymbol{c}$ means adding the constant c to every dimension of x. (5)

2. In the Indian Premier League, the KKR team plays 60 percent of its games at night and 40 percent in the daytime. It wins 55 percent of its night games but only 35 percent of its day games. You read in the paper that the KKR won their last game. What is the probability that it was played at night? (5)

3. Given a set of training example $(\boldsymbol{X}, \boldsymbol{y})$ where $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ are all the input examples and $\boldsymbol{y} \in \mathbb{R}^{m\times 1}$ be the corresponding output. We would like to fit a curve ($y = \boldsymbol{w}^T\boldsymbol{x}$) to predict $y \in \mathbb{R}$ for a given $\boldsymbol{x} \in \mathbb{R}^n$. Derive an expression, involving $\boldsymbol{X}$ and $\boldsymbol{y}$ only, to determine $\boldsymbol{w}$. (5)

4. Let $\boldsymbol{A}$ be a 2×2 matrix with eigen vector $v_1 = [-\frac{1}{\sqrt{2}} \ \frac{1}{\sqrt{2}}]^T$ and $v_2 = [\frac{1}{\sqrt{2}} \ \frac{1}{\sqrt{2}}]^T$ with eigen value of 1 and 3 respectively. We do linear transformation of the points lying on unit circle using matrix $\boldsymbol{A}$. That is, we compute $\boldsymbol{Ax}$ where $\boldsymbol{x}$ lies on unit circle. Draw the plot of $\boldsymbol{Ax}$. (5)

5. Consider the neural network for some classification problem as shown in Figure 1. In the figure, x_1, x_2 are the inputs and y is the final output. Activation functions for h_1, h_2, o_1, o_2 are ReLU. Node y uses softmax for predicting the output. Training examples for the network is as follows:

x_1	x_2	y
1	0.25	1
0.25	1	0

Run back-propagation algorithm using stochastic gradient descent approach with batch size of 2 to determine the new value of w_1 and w_5. Assume cross entropy as the loss function and initial value of $w_1 = 0.1$, $w_2 = 0.2$, $w_3 = 0.3$, $w_4 = -0.4$, $w_5 = 0.5$, $w_6 = 0.6$, $w_7 = 0.7$, $w_8 = 0.8$. (10)

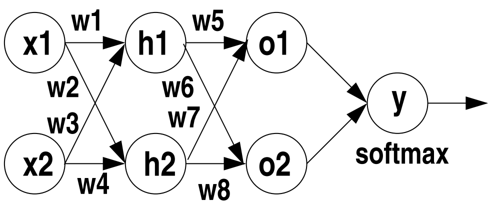

Figure 1: Neural network