Machine Learning Problem Set 3

Travis S. Collier, Graduate Student

14 October 2019

1 Problem 1

Definition of Cross Entropy

$$S(p,q) = -\int_{\mathbb{R}} p(x)\ln q(x)dx \tag{1}$$

1.1 a

Show $S(p_1+p_2,q) = S(p_1,q)+S(p_2,q)$
$S(p_1+p_2,q) = -\int_{\mathbb{R}}(p_1(x)+p_2(x))\ln q(x)dx$ (Applying definition)
$S(p_1+p_2,q) = -\int_{\mathbb{R}} p_1(x)\ln q(x)dx + -\int_{\mathbb{R}} p_1(x)\ln q(x)dx$ (linearity of integrals)
$S(p_1+p_2,q) = S(p_1,q)+S(p_2,q)$ (Applying definition)

1.2 b

Show $S(\alpha p,q) = \alpha S(p,q)$
$S(\alpha p,q) = -\int_{\mathbb{R}} \alpha p(x)\ln q(x)dx$ (Definition)
$S(\alpha p,q) = -\alpha\int_{\mathbb{R}} p(x)\ln q(x)dx$ (Linearity)
$S(\alpha p,q) = \alpha S(p,q)$ (Definition)
Show: $S(\alpha p,q) = S(p,q^\alpha)$
$S(\alpha p,q) = -\alpha\int_{\mathbb{R}} p(x)\ln q(x)dx$
$S(\alpha p,q) = -\int_{\mathbb{R}} p(x)\ln q^\alpha(x)dx$ (Property of ln)
$S(\alpha p,q) = S(p,q^\alpha)$

1.3 b

Show $S(p, q_1q_2) = S(p, q_1) + S(p, q_2)$
$S(p, q_1q_2) = -\int_{\mathbb{R}}(p(x)) \ln(q_1(x)q_2(x))dx$ (Applying definition)
$S(p, q_1q_2) = -\int_{\mathbb{R}}(p(x))(\ln(q_1(x)) + \ln(q_2(x)))dx$ (Property of ln)
$S(p, q_1q_2) = -\int_{\mathbb{R}}(p(x)) \ln(q_1(x))dx + -\int_{\mathbb{R}}(p(x)) \ln(q_2(x))dx$ (Linearity)
$S(p, q_1q_2) = S(p, q_1) + S(p, q_2)$ (defintion)

2 Problem 2

2.1 a

Python code:
from numpy import linspace
def f(x):
return x*x + 1
def g(x):
return x - 0.5

Diff = []
N = [100,1000]
for i in range(len(N)):
Diff.append((i,max([abs(f(x)-g(x)) for x in linspace(0,1,num=N[i])])))
print(Diff)
Diff = [(0,1.5),(1,1.5)]

2.2 b

Using Calculus we calculate:
$f(x) = x^2 + 1$ and $g(x) = x - 0.5$
$h(x) = f(x) - g(x) = x^2 + 1.5 - x$
$h^{'}(x) = 2x - 1$
Solve for 0 to obtain the function maximum:
$2x - 1 = 0 \Rightarrow x = .5$
$h(.5) = 1.25$
$h(0) = 1.5$
$h(1) = 1.5$
$\therefore$ using the classic rules we have found the same maxima

3 Problem 3

The Kullback-Leibler Divergence:

$$D_{KL}(p||q) = -\int_{\mathbb{R}} p(x) \ln \frac{q(x)}{p(x)} dx \tag{2}$$

3.1 a

Given two densities $p_1(x) = \xi^1 e^{-\xi^1 x}$ and $p_2(x) = \xi^2 e^{-\xi^2 x}$ for $x \geq 0$ show
$D_{KL}(p_1||p_2) = \frac{\xi^2}{\xi^1} - \ln \frac{\xi^2}{\xi^1} - 1$
$D_{KL}(p_1||p_2) = -\int_{\mathbb{R}} p_1(x) \ln \frac{p_2(x)}{p_1(x)} dx$
$D_{KL}(p_1||p_2) = -\int_{\mathbb{R}} \xi^1 e^{-\xi^1 x} \ln \frac{\xi^2 e^{-\xi^2 x}}{\xi^1 e^{-\xi^1 x}} dx$
$D_{KL}(p_1||p_2) = -\int_{\mathbb{R}} \xi^1 e^{-\xi^1 x} \ln \frac{\xi^2}{\xi^1} \frac{e^{-\xi^2 x}}{e^{-\xi^1 x}} dx$
$D_{KL}(p_1||p_2) = -\int_{\mathbb{R}} \xi^1 e^{-\xi^1 x} (\ln \frac{\xi^2}{\xi^1} + \ln \frac{e^{-\xi^2 x}}{e^{-\xi^1 x}}) dx$
$D_{KL}(p_1||p_2) = -\int_{\mathbb{R}} \xi^1 e^{-\xi^1 x} \ln \frac{\xi^2}{\xi^1} dx + -\int_{\mathbb{R}} \xi^1 e^{-\xi^1 x} \ln \frac{e^{-\xi^2 x}}{e^{-\xi^1 x}} dx$
$D_{KL}(p_1||p_2) = -\frac{\xi^1}{\xi^1} \ln \frac{\xi^2}{\xi^1} + \int_{\mathbb{R}} \xi^1 e^{-\xi^1 x} \ln \frac{e^{-\xi^2 x}}{e^{-\xi^1 x}} dx$
$D_{KL}(p_1||p_2) = -\frac{\xi^1}{\xi^1} \ln \frac{\xi^2}{\xi^1} + \int_{\mathbb{R}} \xi^1 e^{-\xi^1 x} (\ln e^{-\xi^2 x} - \ln e^{-\xi^1 x}) dx$
$D_{KL}(p_1||p_2) = -\frac{\xi^1}{\xi^1} \ln \frac{\xi^2}{\xi^1} + \int_{\mathbb{R}} \xi^1 e^{-\xi^1 x} (-\xi^2 x + \xi^1 x) dx$
$D_{KL}(p_1||p_2) = -\frac{\xi^1}{\xi^1} \ln \frac{\xi^2}{\xi^1} + \int_{\mathbb{R}} \xi^1 e^{-\xi^1 x} (-\xi^2 + \xi^1) x dx$
$D_{KL}(p_1||p_2) = -\frac{\xi^1}{\xi^1} \ln \frac{\xi^2}{\xi^1} + (-\xi^2 + \xi^1) \xi^1 \int_{\mathbb{R}} x e^{-\xi^1 x} dx$
$D_{KL}(p_1||p_2) = -\frac{\xi^1}{\xi^1} \ln \frac{\xi^2}{\xi^1} + (-\xi^2 + \xi^1) \xi^1 (-\frac{1}{(\xi^1)^2})$
$D_{KL}(p_1||p_2) = -\frac{\xi^1}{\xi^1} \ln \frac{\xi^2}{\xi^1} + (-\xi^2 + \xi^1) \frac{1}{\xi^1}$
$D_{KL}(p_1||p_2) = -\frac{\xi^1}{\xi^1} \ln \frac{\xi^2}{\xi^1} + \frac{\xi^2}{\xi^1} - \frac{\xi^1}{\xi^1}$
$D_{KL}(p_1||p_2) = -\ln \frac{\xi^2}{\xi^1} + \frac{\xi^2}{\xi^1} - 1$

3.2 b

$D_{KL}(p_1||p_2) - D_{KL}(p_2||p_1) = -\ln \frac{\xi^2}{\xi^1} + \frac{\xi^2}{\xi^1} - 1 + \ln \frac{\xi^1}{\xi^2} - \frac{\xi^1}{\xi^2} + 1$
$D_{KL}(p_1||p_2) - D_{KL}(p_2||p_1) = -\ln \frac{\xi^2}{\xi^1} + \frac{\xi^2}{\xi^1} + \ln \frac{\xi^1}{\xi^2} - \frac{\xi^1}{\xi^2}$
$D_{KL}(p_1||p_2) - D_{KL}(p_2||p_1) = \ln \frac{(\xi^1)^2}{(\xi^2)^2} + \frac{(\xi^2)^2 - (\xi^1)^2}{\xi^1 \xi^2}$
This is in general nonzero.

4 Problem 4

The perceptron model of "OR" is: $\sum_i w_i x_i - .5 > 0$ where $w_i = 1, b = -0.5$
The geometric meaning of this function is the line that intersects the boolean square at (0, .5) with constant negative slope, all the values to the right of the line are 1 and all those to the left are 0.

5 Problem 5

5.1 a

One perceptron cannot learn the "XOR" function because it is not linearly separable.

5.2 b

The neural network described by the equation $Y = H(x_1 - x_2 - \frac{1}{2}) + H(x_2 - x_1 - \frac{1}{2})$ can learn the "XOR" function:
$0 = H(0 - 0 - \frac{1}{2}) + H(0 - 0 - \frac{1}{2})$
$1 = H(0 - 1 - \frac{1}{2}) + H(1 - 0 - \frac{1}{2})$
$1 = H(1 - 0 - \frac{1}{2}) + H(0 - 1 - \frac{1}{2})$
$0 = H(1 - 1 - \frac{1}{2}) + H(1 - 1 - \frac{1}{2})$

5.3 c

A three perceptron model can learn "XOR" with the following architecture: Output $= ((x_1 + x_2 + 0.5) + (-x_1 - x_2 - 1.5) + 1.5))$ where the values in the inner parentheses correspond to the two nodes in the hidden layer and the outer parentheses correspond to the output neuron.
This architecture is equivalent to an [["OR","NOT AND"],"AND"] gate.