ICS635 Homework 1: by Lambert Leong

Problem 1

1.

$$\begin{aligned}
E[aX+b] &= \int_{-\infty}^{\infty} (aX+b)f(x)dx \\
&= \int_{-\infty}^{\infty} aXf(x)dx + \int_{-\infty}^{\infty} bf(x)dx \\
&= a\int_{-\infty}^{\infty} Xf(x)dx + b\int_{-\infty}^{\infty} f(x)dx \\
&= aE[X] + b
\end{aligned}$$

2.

$$\begin{aligned}
var(cX) &= E[(cX - c\mu)^2] \\
&= E[cX]^2 - (E[cX])^2 \\
&= c^2E[X^2] - c^2(E[X])^2 \\
&= c^2(E[X^2] - (E[X])^2) \\
&= c^2(E[X] - (E[X])^2) \\
&= c^2 var(X)
\end{aligned}$$

3.

$$\begin{aligned}
var(x) &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \\
&= \int_{-\infty}^{\infty} (x^2 + \mu^2 - 2x\mu)f(x)dx \\
&= \int_{-\infty}^{\infty} x^2 f(x)f(x) + \mu^2 \int_{-\infty}^{\infty} f(x)dx - 2\mu \int_{-\infty}^{\infty} xf(x)dx \\
&= E[X^2] + \mu^2 - 2\mu^2 \\
&= E[X^2] - \mu^2 \\
&= E[X^2] - E[X]^2
\end{aligned}$$

Problem 2

1.

$$\begin{aligned}
E[X] &= \int_{a}^{b} x(\tfrac{1}{b-a})dx \\
&= \tfrac{b+a}{2}
\end{aligned}$$

2.

$$
\begin{aligned}
var(X) &= E[X^2] - E^2[X] \\
&= \int_a^b \tfrac{x^2}{b-a} dx - (\tfrac{b-a}{2})^2 \\
&= \frac{b^2 + ba + a^2}{3} - (\tfrac{b+a}{2})^2 \\
&= \frac{b^2 - 2ba + a^2}{12}
\end{aligned}
$$

Problem 3

1.

2.

3.

Problem 4

1.

$$P(D|T^+) = \tfrac{P(T^+|D)P(D)}{P(T^+)}$$

2.

$$
\begin{aligned}
P(D|T^+) &= \tfrac{P(T^+|D)P(D)}{P(T^+)} \\
&= \tfrac{P(T^+|D)P(D)}{[P(T^+|D) \times P(D)] + [P(T^+|notD) \times P(notD)]} \\
&= \tfrac{SP(D)}{[SP(D)] + [(1-Q)(1-P(D))]} \\
&= \tfrac{.99(.001)}{[.99(.001)] + [(1-.99)(1-.001)]} \\
&= .0902
\end{aligned}
$$

3.

$$
\begin{aligned}
P(D|T^+) &= \tfrac{P(T^+|D)P(D)}{P(T^+)} \\
&= \tfrac{P(T^+|D)P(D)}{[P(T^+|D) \times P(D)] + [P(T^+|notD) \times P(notD)]} \\
&= \tfrac{SP(D)}{[SP(D)] + [(1-Q)(1-P(D))]} \\
&= \tfrac{.99(.001)}{[.99(.001)] + [(1-.90)(1-.001)]} \\
&= .0098
\end{aligned}
$$

Problem 5

1. $\nabla f(x) = \frac{1}{2}[\nabla x^T Ax] + b^T$
   if A symmetric
   $\vec{\nabla}(x^T Ax) = Ax + x^T A = 2Ax$
   Then
   $= \frac{1}{2}(2Ax) + b^T x$
   $= Ax + b^T$

2. chain rule?

3.

4.

Problem 6

1. (a) A small λ would prevent overfiting of the on the training set and lead to an increase in error.
   (b) The model may thus be able to generalize well to the validation set and increase error.
   (c) w will increase
   (d) The number of non-zero elemnts in w would also increase

2. (a) A big λ could lead to over fitting on the training set which would minimize error
   (b) Overfitting on the training set would lead to an increased error during validation
   (c) w or the bias will be low as well
   (d) There will be few non-zero elements of w.

3. (a) An L2 regularization would increase the number of non-zero elements.
   (b)
   (c)

Problem 7

1. The outcomes of the XOR functions are not linearly sperable. Therefore, no parameters exists that can lead to a zero loss for a single neuron.

2. h1 = np.heaviside((1*x1+1*x2+(-0.5)),.5)
   h2 = np.heaviside(((-1)*x1+(-1)*x2+(1.5)),.5)
   y = np.heaviside((1*(h1)+(1)*(h2)+(-1.5)),.5)

Problem 8

1.

2.

3.