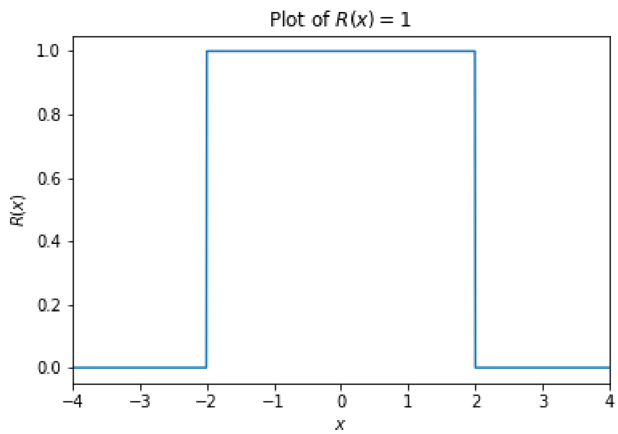
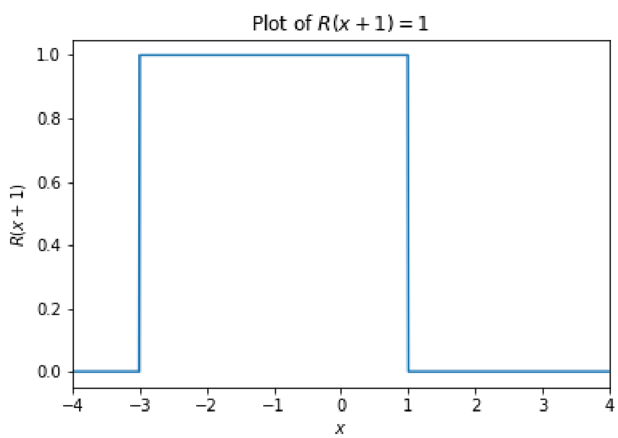
ESE 541 Week 5 & 6 Practice Problems - Solutions

Problem 1. Define the rectangular function as $R(x) = 1$ for $x \in [-2, 2]$, 0 otherwise:

1. Plot $R(x)$.



2. Plot $R(x + 1)$.



The $+1$ term results in a left shift of the function.
If we plug in $x = -3$ into $R(x + 1)$, we get $R(-2)$, which is defined to be 1 as can be seen in the plot in (1).

Problem 2. Define the sigmoid function as $S(x) = \frac{1}{1+e^{-x}}$:

1. Show that $S(x)$ can also be written as $\frac{e^x}{e^x+1}$.

$$\begin{aligned} S(x) &= \frac{1}{1+e^{-x}} && \text{rewrite given equation} \\ &= \frac{1}{1+e^{-x}} \cdot \frac{e^x}{e^x} && \text{multiply by } \frac{e^x}{e^x} \\ &= \frac{e^x}{e^x+1} && \text{use } e^{-x} \cdot e^x = 1 \end{aligned} \tag{1}$$

2. Show that $S(-x)$ can be written as $\frac{1}{1+e^{x}}$ and $\frac{e^{-x}}{e^{-x}+1}$

$$\begin{aligned}
S(-x) &= \frac{1}{1+e^{-(-x)}} && \text{rewrite given equation plugging in } -x \\
&= \frac{1}{1+e^{x}} && \text{simplify} \\
&= \frac{1}{1+e^{x}} \cdot \frac{e^{-x}}{e^{-x}} && \text{multiply by } \frac{e^{-x}}{e^{-x}} \\
&= \frac{e^{-x}}{e^{-x}+1} && \text{use } e^{x} \cdot e^{-x} = 1
\end{aligned} \tag{2}$$

3. Find an expression for $1 - S(x)$. What is the relationship between $1 - S(x)$ and $S(-x)$?

$$\begin{aligned}
1 - S(x) &= 1 - \frac{1}{1+e^{-x}} && \text{rewrite given equation} \\
&= \frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}} && \text{find a common denominator} \\
&= \frac{1+e^{-x}-1}{1+e^{-x}} && \text{subtract the fractions} \\
&= \frac{e^{-x}}{e^{-x}+1} && \text{does this look familar?}
\end{aligned} \tag{3}$$

We find that $1 - S(x)$ and $S(-x)$ are the same

Problem 3. Let $P(Y = +1) = P(Y = -1) = 0.5$. We are also given:

$$P(X = x|Y = +1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-3)^2}{2}} \qquad P(X = x|Y = -1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+3)^2}{2}}$$

These conditional probabilities can also be expressed as $P(x| + 1) \sim \mathcal{N}(3, 1)$ and $P(x| - 1) \sim \mathcal{N}(-3, 1)$, where $\mathcal{N}(\mu, \sigma^2)$ represents a normal distribution centered at mean μ with standard deviation σ.

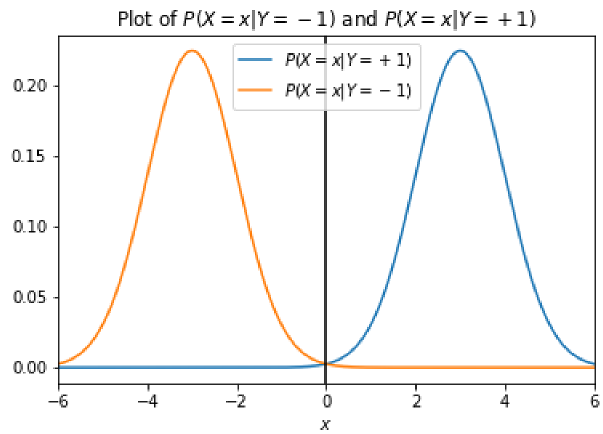
1. Find an expression for $P(X = x, Y = y)$

$$\begin{aligned}
P(X = x, Y = y) &= P(X = x|Y = y) \cdot P(Y = y) && \text{using definition of conditional probability} \\
\\
P(X = x, Y = +1) &= P(X = x|Y = +1) \cdot P(Y = +1) && \text{solve joint distribution for } Y = +1 \text{ case} \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-3)^2}{2}} \cdot \frac{1}{2} \\
&= \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-3)^2}{2}} \\
P(X = x, Y = -1) &= P(X = x|Y = -1) \cdot P(Y = -1) && \text{solve joint distribution for } Y = -1 \text{ case} \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+3)^2}{2}} \cdot \frac{1}{2} \\
&= \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x+3)^2}{2}}
\end{aligned} \tag{4}$$

Because the only difference between the cases is the sign of the $(x \pm 1)$ term, we can introduce $y \in \{-1, 1\}$ to create a general equation

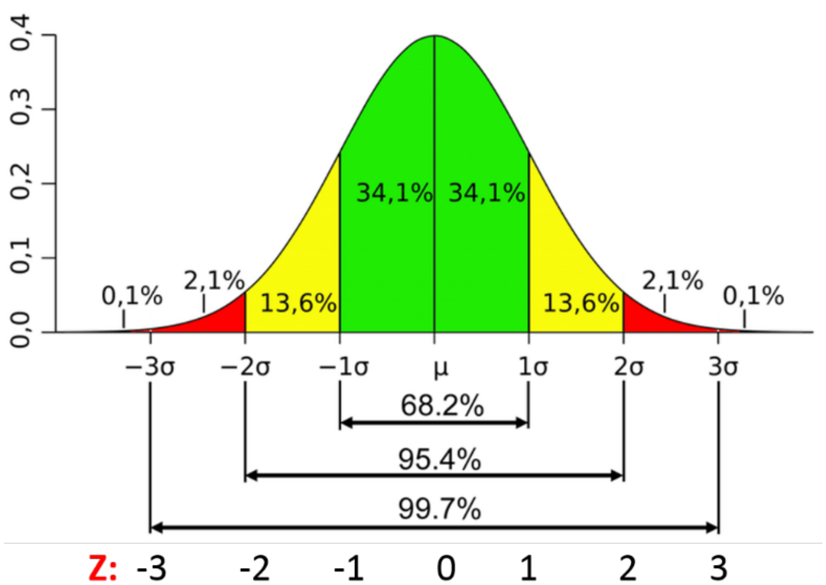
$$P(X = x, Y = y) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-3y)^2}{2}} \tag{5}$$

2. Plot the conditional distributions $P(X = x|Y = +1)$ and $P(X = x|Y = -1)$. On the graph, show the Bayes optimal classification rule.



If $x < 0$, the Bayes optimal classifier will yield $y = -1$
If $x > 0$, the Bayes optimal classifier will yield $y = +1$

3. Calculate the mass of $P(X = x|Y = -1)$ that lies on the wrong side of the decision rule



For a normal distribution, the total mass of the curve must be 1 (by the requirements of a probability density function). We also know that 99.7% of its mass is contained within 3 standard deviations. Because $P(X = x|Y = -1)$ is centered at $\mu = -3$ with $\sigma = 1$, the location of the decision rule at $x = 0$ is 3 standard deviations away. Since we only care about mass to the right of 0, and not to the left of -6 (3 standard deviations away to the left of the mean), we must divide the remaining area outside of 3 standard deviations by 2. Thus:

$$\text{Total mass on the wrong side of the decision rule} = \frac{1 - 0.997}{2} = 0.0015$$

Problem 4. Explain what the following represent in the context of detecting spam email

1. False Positive Rate (FPR)

   The rate at which an email that is not actually spam is marked by your model as spam

2. False Negative Rate (FNR)

   The rate at which an email that is actually spam is marked by your model as being not spam

3. True Positive Rate (TPR)

   The rate at which an email that is actually spam is marked by your model as being spam

Problem 5. Suppose we have a dataset of 100 students for which we have access to two features: X_1='hours studied', X_2='GPA', and Y = 'receive an A'. We fit a logistic regression and produce the following estimated coefficients: $\hat{\beta}_0 = -6$, $\hat{\beta}_1 = 0.05$, and $\hat{\beta}_2 = 1$.

1. Estimate the probability that a student who studies for 40 hours and has a GPA of 3.5 gets an A in the class.

We use logistic regression with the following sigmoid:

$$p(X) = \frac{e^{\hat{\beta}_0+\hat{\beta}_1 X_1+\hat{\beta}_2 X_2}}{1+e^{\hat{\beta}_0+\hat{\beta}_1 X_1+\hat{\beta}_2 X_2}} = \frac{e^{-6+0.05(40)+1(3.5)}}{1+e^{-6+0.05(40)+1(3.5)}} \approx 0.378$$

2. How many hours would the student in part (a) need to study to have a 50% chance of getting an A in the class?

Logistic regression can be written using the log odds (logit) transformation of $p(X)$. Since we want a 50% chance of getting an A, we take $p(X) = 0.50$.

$$\ln\left(\frac{p(X)}{1-p(X)}\right) = \ln\left(\frac{0.50}{1-0.50}\right) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$$

$$\to \ln(1) = 0 = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$$

$$\to 0 = -6 + 0.05(X_1) + 1(3.5)$$

$$\therefore X_1 = 50$$

The student would have to need to study 50 hrs to have a 50% chance of getting an A in the class.

Problem 6. Prove that

$$\log\left(\frac{p\,(X)}{1-p\,(X)}\right) = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p$$

is equivalent to

$$p\,(X) = \frac{\exp\left(\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p\right)}{1+\exp\left(\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p\right)}.$$

Steps for the proof are provided in one of the quizzes. Note that an alumnus actually got this as an interview question!

Problem 7: Consider a communication system transmitting a variable Y that can take two possible values, either 0 or 1. The communication channel is noisy, so that the receptor receives a noisy version of Y equal to $X = 2Y + \varepsilon$, where ε is a Gaussian $\mathcal{N}\left(-1, \frac{1}{4}\right)$. We also know that the communication system transmits, in expectation, 50% of 0's and 50% of 1's.

1. Write down expressions for $f_0(x) = \Pr(X = x|Y = 0)$ and $f_1(x) = \Pr(X = x|Y = 1)$? What are the values of the priors, $\pi_0 = \Pr(Y = 0)$ and $\pi_1 = \Pr(Y = 1)$?

$$
\begin{aligned}
f_0(x) &= \frac{1}{\sqrt{2\pi(\frac{1}{4})}} e^{-\frac{(x+1)^2}{2(\frac{1}{4})}} = \sqrt{\frac{2}{\pi}} e^{-2(x+1)^2} \\
f_1(x) &= \frac{1}{\sqrt{2\pi(\frac{1}{4})}} e^{-\frac{(x-1)^2}{2(\frac{1}{4})}} = \sqrt{\frac{2}{\pi}} e^{-2(x-1)^2} \\
\pi_0 &= \frac{1}{2} \\
\pi_0 &= \frac{1}{2}
\end{aligned}
$$

2. Write down the classifier $\hat{Y} = f(x)$ corresponding to a linear discriminant analysis.

$$
\begin{aligned}
\hat{Y} &= \underset{k\in\{0,1\}}{\arg\max}\, f_k(x)\pi_k \\
&= \underset{k\in\{0,1\}}{\arg\max} \sqrt{\frac{2}{\pi}} e^{-2(x+1-2k)^2} \cdot \frac{1}{2}
\end{aligned}
$$

We have two Gaussians with identical variances, weighted by identical priors. Hence, they will intersect at the midpoint of their means. We find the intersection point occurs at $x = 0$. If $x \geq 0$, classify as 1. Else, classify as 0.

3. Compute the false positive rate and true positive rate of the classifier.

A false positive occurs when the instance's true value is 0, but we misclassify as 1. Thus, we need to compute the area of $f_0(x)$ that is to the right of 0.

0 is 2 standard deviations from the mean of -1. As can be seen above in problem 3.3, 95.4% of the mass is contained within 2 standard deviations. Thus, with probability $1 - 0.954 = 0.046$, an entry will fall outside 2 standard deviations. We only care about those to the right of 0, so we can divide by two (based on symmetry). Hence, for all true 0 instances, with probability $\frac{.046}{2} = 0.023$, our classifier will classify 1. We also know that it'll be classified as a true negative with probability $1 - 0.023 = 0.977$

To find the probability of a false negative (where the true instance is of value 1, but we label 0), we can appeal to symmetry since $f_1(x)$ has the same distribution, just with mean 1. Thus, the probability a generic instance is misclassified as a false negative is also 0.023 and the probability it is classified as a true positive is $1 - 0.023 = 0.977$.

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{0.977}{0.977 + 0.023} = 0.977$$

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} = \frac{0.023}{0.977 + 0.023} = 0.023$$

4. What is the total error rate of the classifier?

$$\text{Error Rate} = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{0.023 + 0.023}{0.977 + 0.977 + 0.023 + 0.023} = 0.023$$