# CS534 — Written Homework Assignment 0 —*Solution*

This assignment is to get you refreshed on mathematical concepts that are important for machine learning. You can type your solution out in the provided latex file. Recommended if you are fluent (or want to get fluent) in math mode of latex. Writing out the solution by hand and submitting a scanned pdf is also acceptable — just be sure that your writing is clear and easily readable.

## Linear algebra

1. (Let's start with some basic matrix multiplications) Compute the following matrix products if possible:

   (a) (1pt)

   $$\begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

   *Dimensions don't match. Impossible.*

   (b) (1pt)

   $$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

   $\begin{bmatrix} 4 & 3 & 5 \\ 10 & 9 & 11 \\ 16 & 15 & 17 \end{bmatrix}$

   (c) (1pt)

   $$\begin{bmatrix} 1 & 2 & 1 & 2 \\ 4 & 1 & -1 & -4 \end{bmatrix} \begin{bmatrix} 0 & 3 \\ 1 & -1 \\ 2 & 1 \\ 5 & 2 \end{bmatrix}$$

   $\begin{bmatrix} 14 & 6 \\ -21 & 2 \end{bmatrix}$

   (d) (1pt)

   We have $\mathbf{x} = [1, 0, 1]^T$, answer following questions.

   - What is the norm of $\mathbf{x}$?
   - What is $\mathbf{x}^T\mathbf{x}$?
   - What is $\mathbf{x}\mathbf{x}^T$?

   *The norm of $\mathbf{x}$ is $\sqrt{2}$, and $\mathbf{x}^T\mathbf{x}$ is simply the squared norm of $\mathbf{x}$: 2*

   $$\mathbf{x}\mathbf{x}^T = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

2. (let's get comfortable with notations). Consider the following set of linear equations:

$$\begin{cases} x_1 + x_2 - x_3 - x_4 = 1 \\ 2x_1 + 5x_2 - 7x_3 - 5x_4 = -2 \\ 2x_1 - x_2 + x_3 + 3x_4 = 4 \\ 5x_1 + 2x_2 - 4x_3 - 2x_4 = 6 \end{cases}$$

(a) (1 pt) Please express the system of equations as $A\mathbf{x} = \mathbf{b}$ by specifying the matrix $A$ and vector $\mathbf{b}$

$$A = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 2 & 5 & -7 & -5 \\ 2 & -1 & 1 & 3 \\ 5 & 2 & -4 & -2 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ -2 \\ 4 \\ 6 \end{bmatrix}$$

(b) (1 pt) Solve for $A\mathbf{x} = \mathbf{b}$ by using the matrix inverse of $A$ (you can use software to compute the inverse).

$$A^{-1} = \begin{bmatrix} 1/2 & -1/6 & 0 & 1/6 \\ 2 & 1/6 & 1/2 & -2/3 \\ 7/4 & -1/4 & 0 & -1/4 \\ -1/4 & 1/4 & 1/2 & -1/4 \end{bmatrix}$$

$$\mathbf{x} = \mathbf{A^{-1}b} = \begin{bmatrix} 11/6 \\ -1/3 \\ 3/4 \\ -1/4 \end{bmatrix}$$

# Vector Calculus

1. (start with simple derivatives). Compute the derivative $f'(x)$ for

(a) (1 pts) the logistic (aka sigmoid) function $f(x) = \frac{1}{1+\exp(-x)}$

$$\begin{aligned} f'(x) &= \frac{-1}{(1+exp(-x))^2} \times \frac{d(1+exp(-x))}{dx} \\ &= \frac{-1}{(1+exp(-x))^2} \times \frac{d(exp(-x))}{dx} \\ &= \frac{-1}{(1+exp(-x))^2} \times (-exp(-x)) = \frac{exp(-x)}{(1+exp(-x))^2} \\ &= \frac{1}{(1+exp(-x))} \times \frac{exp(-x)}{(1+exp(-x))} = f(x) \times (1 - f(x)) \end{aligned}$$

(b) (1 pts) $f(x) = \exp(-\frac{1}{2\sigma^2}(x-\mu)^2)$

$$\begin{aligned} f'(x) &= \exp(-\frac{1}{2\sigma^2}(x-\mu)^2) \times \frac{d(-\frac{1}{2\sigma^2}(x-\mu)^2)}{dx} \\ &= \exp(-\frac{1}{2\sigma^2}(x-\mu)^2) \times (-\frac{2}{2\sigma^2}(x-\mu)) \\ &= \frac{-(x-\mu)}{\sigma^2} exp(-\frac{(x-\mu)^2}{2\sigma^2}) \end{aligned}$$

2. (work out gradients). Compute the gradient $\nabla_{\mathbf{x}} f$ of the following functions. Please clearly specify the dimension of the gradient.

   (a) (1pt)

$$f(z) = \log(1+z),\ z = \mathbf{x}^T\mathbf{x},\ \mathbf{x} \in R^D$$

   *Use the property for derivative of a log first, then apply the chain rule and vector derivative properties.*

   $\nabla_{\mathbf{x}} f = \frac{1}{1+z} * \nabla_{\mathbf{x}} z$, *where* $z = \mathbf{x}^{\mathbf{T}}\mathbf{x}$.

   *Now, we have* $\nabla_{\mathbf{x}} z = \nabla_{\mathbf{x}} \mathbf{x}^{\mathbf{T}}\mathbf{x} = 2\mathbf{x}$ *(see the Matrix derivative cheatsheet or Matrix cookbook). Finally, sub in* $\mathbf{x}^{\mathbf{T}}\mathbf{x}$ *for* $z$*:*
   $\nabla_{\mathbf{x}} f = \frac{2\mathbf{x}}{1+\mathbf{x}^{\mathbf{T}}\mathbf{x}} \in R^D$

   (b) (2pts)

$$f(z) = \exp\left(-\frac{1}{2}z\right)$$
$$z = g(\mathbf{y}) = \mathbf{y}^T S^{-1}\mathbf{y}$$
$$\mathbf{y} = h(\mathbf{x}) = \mathbf{x} - \mu$$

where $\mathbf{x}, \mu \in R^D, S \in R^{D\times D}$ is a symmetric matrix.

   (Consult the matrix cookbook section 2.4)
   *Recall first the property (from Matrix cookbook) that:*
   $\nabla_{\mathbf{x}} \mathbf{x}^{\mathbf{T}} M\mathbf{x} = (M+M^T)\mathbf{x}$, *and* $2M\mathbf{x}$ *if* $M$ *is symmetric. Here we have* $M = S^{-1}$ *and is a symmetric matrix. The rest comes directly from derivatives of functions and the chain rule:*

   $\nabla_{\mathbf{x}} f = -\frac{1}{2}\exp^{-\frac{1}{2}z}(\nabla_{\mathbf{x}} z)$
   $\nabla_{\mathbf{x}} f = -\frac{1}{2}\exp^{-\frac{1}{2}z} 2S^{-1}\mathbf{y}(\nabla_{\mathbf{x}}\mathbf{y})$ *(from statement above)*
   $\nabla_{\mathbf{x}} f = -\frac{1}{2}\exp^{-\frac{1}{2}z} 2S^{-1}\mathbf{y}\nabla_{\mathbf{x}}(\mathbf{x}-\mu)$ *(note* $\nabla_{\mathbf{x}}(\mathbf{x}-\mu) = I$*)*
   $\nabla_{\mathbf{x}} f = -\frac{1}{2}\exp^{-\frac{1}{2}z} *2S^{-1}\mathbf{y}$ *(now, substitute in the* $\mathbf{x}-\mu$*)*
   $\nabla_{\mathbf{x}} f = -\frac{1}{2}\exp^{-\frac{1}{2}(\mathbf{x}-\mu)^T S^{-1}(\mathbf{x}-\mu)} *2S^{-1}(\mathbf{x}-\mu) \in R^D$

## Probability

1. Consider two discrete random variables $X$ and $Y$ with the following joint distribution:

| $Y$ | | | | | |
|---|---|---|---|---|---|
| $y_1$ | 0.01 | 0.02 | 0.03 | 0.1 | 0.1 |
| $y_2$ | 0.05 | 0.1 | 0.05 | 0.07 | 0.2 |
| $y_3$ | 0.1 | 0.05 | 0.03 | 0.05 | 0.04 |
| $X$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |

   Please compute:

(a) (1 pt) The marginal distributions $p(x)$ and $p(y)$

$p(y)$ :

| $y_1$ | $y_2$ | $y_3$ |
|---|---|---|
| 0.26 | 0.47 | 0.27 |

$p(x)$ :

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|---|---|---|---|---|
| 0.16 | 0.17 | 0.11 | 0.22 | 0.34 |

(b) (1 pt) The Conditional distribution $p(x|Y = y_1)$ and $p(y|X = x_3)$

$p(x|Y = y_1)$:

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|---|---|---|---|---|
| 1/26 | 2/26 | 3/26 | 10/26 | 10/26 |

$p(y|X = x_3)$:

| $y_1$ | $y_2$ | $y_3$ |
|---|---|---|
| 3/11 | 5/11 | 3/11 |

2. Consider two coins, one is fair and the other one has a 1/10 probability for head. Now you randomly pick one of the coins, and toss it twice. Answer the following questions.

(a) (1pt) What is the probability that you picked the fair coin?
*0.5*

(b) (1pt) What is the probability of the first toss being head?
*Let $x_1$ denote the outcome of the first toss and let $y$ denote the coin that is selected. We can write down the following probabilities.*
$P(y = f) = P(y = u) = \frac{1}{2}$
$P(x_1 = h|y = f) = \frac{1}{2}$
$p(x_1 = h|y = u) = 1/10$

*Now we can write out the probability of the first toss being head as:*
$P(x_1 = h) = P(x_1 = h, y = f) + P(x_1 = h, y = u)$
$= P(y = f)P(x_1 = h|y = f) + P(y = u)P(x_1 = h|y = u) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{10}$
$= \frac{1}{2} \times \frac{6}{10} = \frac{3}{10}$

(c) (2pts) If both tosses are heads, what is the probability that you have chosen the fair coin (Hint: you should apply Bayes Rule for this)?
*Let $x_1$, $x_2$ denote the outputs of the first two tosses. It is easy to see that*
$P(x_1 = h, x_2 = h|y = f) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$
$P(x_1 = h, x_2 = h|y = u) = \frac{1}{10} \times \frac{1}{10} = \frac{1}{100}$

*Now we need to compute $P(y = f|x_1 = h, x_2 = h)$, to do so, we use Bayes Theorem:*
$P(y = f|x_1 = h, x_2 = h) = \frac{P(x_1=h,x_2=h|y=f)P(y=f)}{P(x_1=h,x_2=h)}$
*To compute the denominator, we use the same approach as used in (a):*
$P(x_1 = h, x_2 = h) = P(x_1 = h, x_2 = h|y = f)P(y = f) + P(x_1 = h, x_2 = h|y = u)P(y = u)$
$= \frac{1}{4} \times \frac{1}{2} + \frac{1}{100} \times \frac{1}{2} = \frac{13}{100}$

*Plug this into the Bayes Theorem, we have:*
$P(y = f|x_1 = h, x_2 = h) = \frac{P(x_1=h,x_2=h|y=f)P(y=f)}{P(x_1=h,x_2=h)} = \frac{1/4\times 1/2}{13/100} = \frac{25}{26}$

(d) (2pts) If both tosses are heads, what is the probability that the third coin toss will be head? (you should build on results of c)
*We will use $x_3$ to denote the outcome of the third coin toss.*

$$\begin{aligned}
&P(x_3 = h|x_1 = h, x_2 = h)\\
&= P(x_3 = h, y = f|x_1 = h, x_2 = h) + P(x_3 = h, y = u|x_1 = h, x_2 = h)\\
&= P(x_3 = 1|y = f)P(y = f|x_1, x_2) + P(x_3 = 1|y = u)P(y = u|x_1, x_2)\\
&= \frac{1}{2} \times \frac{25}{26} + \frac{1}{10} \times \frac{1}{26}\\
&= \frac{63}{130}
\end{aligned}$$