Discussion 1

Note: *Your TA will probably not cover all the problems on this worksheet. The discussion worksheets are not designed to be finished within an hour. They are deliberately made slightly longer so they can serve as resources you can use to practice, reinforce, and build upon concepts discussed in lectures, discussions, and homework.*

Chain Rule:

$$\frac{\partial}{\partial x} f(g(x) + h(x)) = \frac{\partial f}{\partial g}\frac{dg}{dx} + \frac{\partial f}{\partial h}\frac{dh}{dx} \tag{1}$$

1 Calculus

(a) Consider the function

$$f(x, y) = \sigma(ax + by),$$

where $a, b \in \mathbb{R}$ and $\sigma(t) = \dfrac{1}{1 + e^{-t}}$ for $t \in \mathbb{R}$.

(i) Show that $\dfrac{d\sigma}{dt} = \sigma(t)\big(1 - \sigma(t)\big)$.

(ii) Using the result you showed in part (i) and the chain rule, compute $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$.

Solution:

(i)

$$\begin{aligned}
\frac{d}{dt}\sigma(t) &= \frac{d}{dt}\left(\frac{1}{1 + e^{-t}}\right) && (2)\\
&= \frac{e^{-t}}{(1 + e^{-t})^2} && (3)\\
&= \left(\frac{1}{1 + e^{-t}}\right)\left(\frac{1 + e^{-t}}{1 + e^{-t}} - \frac{1}{1 + e^{-t}}\right) && (4)\\
&= \boxed{\sigma(t)\big(1 - \sigma(t)\big)} && (5)
\end{aligned}$$

(ii) First let's define a new variable $u = ax + by$ such that $f = \sigma(u)$. By chain rule, we have

$$\frac{\partial f}{\partial x} = \frac{\partial \sigma}{\partial u}\frac{\partial u}{\partial x}.$$

We can calculate $\frac{\partial u}{\partial x} = a$, giving us, in conjunction with the result from part (i),

$$\frac{\partial f}{\partial x} = \sigma(u)\left(1 - \sigma(u)\right) \cdot a.$$

Substituting $ax + by$ back in for u gives us

$$\frac{\partial f}{\partial x} = \boxed{a\sigma(ax + by)\left[1 - \sigma(ax + by)\right].}$$

Following the same process for $\frac{\partial f}{\partial y}$, we get

$$\frac{\partial f}{\partial y} = \boxed{b\sigma(ax + by)\left[1 - \sigma(ax + by)\right].}$$

(b) For $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$, define

$$r(\mathbf{x}) = \sum_{j=1}^{n} x_j^2.$$

Compute the partial derivative $\frac{\partial r}{\partial x_i}$ for a generic coordinate $i \in \{1, ..., n\}$.

Solution:

Since $r(\mathbf{x}) = \sum_{j=1}^{n} x_j^2$, we have

$$\frac{\partial r}{\partial x_i} = \boxed{2x_i.}$$

(c) Let $\mathbf{w} \in \mathbb{R}^n$ be a constant vector and define the scalar function

$$s(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} = \sum_{j=1}^{n} w_j x_j.$$

Compute $\frac{\partial s}{\partial x_i}$ for a generic coordinate $i \in \{1, ..., n\}$.

Solution:

With $s(\mathbf{x}) = \sum_{j=1}^{n} w_j x_j$, treating $\mathbf{w}$ as constant,

$$\frac{\partial s}{\partial x_i} = \boxed{w_i.}$$

2 Linear Algebra

(a) Prove that $\mathbf{A}^\top\mathbf{A}$ is symmetric for any $\mathbf{A} \in \mathbb{R}^{m\times n}$.

Solution:

For a matrix to be symmetric, the following identity must hold: $\mathbf{S} = \mathbf{S}^\top$

$$\boxed{(\mathbf{A}^\top\mathbf{A})^\top = \mathbf{A}^\top(\mathbf{A}^\top)^\top = \mathbf{A}^\top\mathbf{A}.}$$

(b) Consider the matrix

$$\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 2 & 1 \end{bmatrix}$$

Find the singular values of $\mathbf{B}$.

Solution:

The singular values are the square roots of the eigenvalues of $\mathbf{B}^\top\mathbf{B}$. First compute:

$$\mathbf{B}^\top\mathbf{B} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 2 & 5 \end{bmatrix}.$$

The characteristic polynomial is:

$$\det\left(\begin{bmatrix} 5 & 2 \\ 2 & 5 \end{bmatrix} - \lambda I\right) = (5-\lambda)^2 - 4 = \lambda^2 - 10\lambda + 21.$$

This factors as $(\lambda - 7)(\lambda - 3)$. The eigenvalues are the roots of this expression, so they are $\lambda_1 = 7$ and $\lambda_2 = 3$. Therefore, the singular values are:

$$\boxed{\sigma_1 = \sqrt{7}, \quad \sigma_2 = \sqrt{3}.}$$

3 Probability Review

An incoming email is spam with prior $p(S) = 0.2$ and not spam with $p(\bar{S}) = 0.8$. Two independent filters flag spam:

$$p(F_1{=}1 \mid S) = 0.9, \quad p(F_1{=}1 \mid \bar{S}) = 0.1, \qquad p(F_2{=}1 \mid S) = 0.8, \quad p(F_2{=}1 \mid \bar{S}) = 0.05,$$

and F_1, F_2 are independent *given* the class (S or $\bar{S}$).

(a) What is the probability that both filters flag an email as spam?

Solution:

We want to find $p(F_1 = 1, F_2 = 1)$. Conditional independence tells us

$$p(F_1 = 1, F_2 = 1 \mid c) = p(F_1 = 1 \mid c)\, p(F_2 = 1 \mid c).$$

We can then use the definition of conditional probability to find

$$p(F_1 = 1, F_2 = 1, c) = p(F_1 = 1, F_2 = 1 \mid c)\, p(c).$$

Finally, we can marginalize over c using the sum rule to get

$$p(F_1 = 1, F_2 = 1) = \sum_{c \in \{S, \bar{S}\}} p(F_1 = 1, F_2 = 1, c).$$

Putting this all together gives us

$$\begin{aligned}
p(F_1{=}1, F_2{=}1) &= \sum_{c \in \{S, \bar{S}\}} p(c)\, p(F_1{=}1 \mid c)\, p(F_2{=}1 \mid c) && (6)\\
&= 0.2 \cdot (0.9 \cdot 0.8) + 0.8 \cdot (0.1 \cdot 0.05) && (7)\\
&= 0.144 + 0.004 && (8)\\
&= \boxed{0.148} && (9)
\end{aligned}$$

(b) Given that both filters flag an email, what is the probability of the email being spam? (You can leave your answer as an unsimplified fraction.)

Solution:

We wish to calculate $p(S \mid F_1 = 1, F_2 = 1)$. We can utilize Bayes' Theorem, which gives us

$$p(S \mid F_1 = 1, F_2 = 1) = \frac{p(F_1 = 1, F_2 = 1 \mid S)\, p(S)}{p(F_1 = 1, F_2 = 1)}.$$

The denominator comes from part (a), $p(S)$ is given to us as 0.2, and $p(F_1 = 1, F_2 = 1 \mid S) = p(F_1 = 1 \mid S)\, p(F_2 = 1 \mid S) = 0.9(0.8) = 0.72$ by conditional independence. Therefore, we have

$$p(S \mid F_1 = 1, F_2 = 1) = \frac{0.72(0.2)}{0.148} = \boxed{\frac{0.144}{0.148} \approx 0.973.}$$