Your Name: 

Your NetID: 

Instructions

• Please write your name on the top of every page.

• This will be a CLOSED BOOK, CLOSED NOTES exam. You are permitted to bring and use only one 8.5x11 page of hand-written notes, front and back.

• No electronic devices (phones, tablets, calculators, computers etc.) are allowed.

• No calculators are permitted. You need not simplify explicit numerical expressions.
Possibly Useful Formulas

**Probability:** \( P(B = 1 | A = 1) = \frac{P(A = 1, B = 1)}{P(A = 1)} \)

**Naïve Bayes:** \( P(X = x | Y = y) \approx \prod_{i=1}^{n} P(W = w_i | Y = y) \)

**Laplace Smoothing:** \( P(w) = \frac{\text{Count}(w) + k}{\sum_w \text{Count}(w) + k(1 + \sum_w 1)} \)

**Perceptron:** \( \vec{w}_y = \vec{w}_y + \eta \vec{x}, \quad \vec{w}_f(x) = \vec{w}_f(x) - \eta \vec{x} \)

**Linear Regression w/SGD:** \( \vec{w} \leftarrow \vec{w} - \frac{\eta}{2} \nabla_{\vec{w}} \epsilon_i^2 = \vec{w} - \eta \epsilon_i \vec{x}_i \)

**Logistic Regression:** \( \nabla_{\vec{w}_c} \mathcal{L}_i = \nabla_{\vec{w}_c}\left(-\ln \frac{e^{\vec{w}_c \vec{y}_i}}{\sum_k e^{\vec{w}_k \vec{y}_i}}\right) = \left(\frac{e^{\vec{w}_c \vec{y}_i}}{\sum_k e^{\vec{w}_k \vec{y}_i}} - y_i \right) \vec{x}_i \)

**Neural Net:** \( \xi_j^{(l)} = b_j^{(l)} + \sum_k w_{jk}^{(l)} h_k^{(l-1)}, \quad h_j^{(l)} = g^{(l)} \left( \xi_j^{(l)} \right) \)

**Back-Propagation:** \( \frac{\partial \mathcal{L}}{\partial h_k^{(l-1)}} = \sum_j \frac{\partial \mathcal{L}}{\partial h_j^{(l)}} \frac{\partial h_j^{(l)}}{\partial h_k^{(l-1)}} \)

**Pinhole Camera:** \( \frac{x'}{f} = -\frac{x}{z}, \quad \frac{y'}{f} = -\frac{y}{z} \)
Question 1  (7 points)  
Consider two binary random variables, $X$ and $Y$. Suppose that

$$P(X = 1) = a$$
$$P(Y = 1) = b$$
$$P(X = 1, Y = 0) = c$$

In terms of $a$, $b$, and/or $c$, what is $P(Y = 1|X = 1)$?
**Question 2  (7 points)**

You’ve been asked to create a naïve Bayes model of the candy produced by the Santa Claus Candy Company. As your training dataset, you’ve been given a box containing 80 pieces of candy, of which 8 are strawberry, 48 are raspberry, and 24 are blueberry. In terms of the Laplace smoothing parameter $k$, estimate the following probabilities:

\[
\begin{align*}
P(\text{flavor=strawberry} | \text{Santa Claus Candy Company}) &= \\
P(\text{flavor=raspberry} | \text{Santa Claus Candy Company}) &= \\
P(\text{flavor=blueberry} | \text{Santa Claus Candy Company}) &= \\
P(\text{flavor=other} | \text{Santa Claus Candy Company}) &= 
\end{align*}
\]
Question 3  (7 points)  
Describe, in one sentence each, the purpose of (1) a training set, (2) a development test set, (3) an evaluation test set.
Question 4  (7 points)
You’re trying to create a multi-class perceptron that will classify animals as being either fish, birds, or reptiles. Your feature vector is $\mathbf{x} = [x_1, x_2, x_3, 1]^T$, where

$$
\begin{align*}
x_1 &= \text{fraction of time the animal spends under water} \\
x_2 &= \text{fraction of time the animal spends on land} \\
x_3 &= \text{fraction of time the animal spends flying}
\end{align*}
$$

- Based on your extensive prior knowledge of zoology, you initialize your perceptron with the following weight vectors: $\mathbf{w}_{\text{fish}} = [1, 0, 0, 0]^T$, $\mathbf{w}_{\text{reptile}} = [0, 1, 0, 0]^T$, and $\mathbf{w}_{\text{bird}} = [0, 0, 1, 0]^T$.
- Your first training token is a crocodile, for which $y = \text{reptile}$, and $\mathbf{x} = [0.7, 0.3, 0, 1]^T$.

After training with this training token, what are the numerical values of $\mathbf{w}_{\text{fish}}$, $\mathbf{w}_{\text{reptile}}$, and $\mathbf{w}_{\text{bird}}$? Assume a learning rate of $\eta = 1$. 
Question 5  (7 points) In stochastic gradient descent, we train using one training token at a time. Suppose
\[ \mathcal{L} = (\vec{w}^T \vec{x} - y)^2 \]
\[ \vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ b \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} \]

In terms of \( \vec{x}, \vec{w}, w_1, w_2, b, x_1, x_2, \) and/or \( y, \) what is \( \frac{d\mathcal{L}}{dw_2}? \)
Question 6  (7 points) 

Suppose that

\[ f = w_1^{(2)} h_1 + w_2^{(2)} h_2 + b^{(2)} \]

\[ h_1 = \text{ReLU} \left( w_1^{(1)} x_1 + w_2^{(1)} x_2 + b_1^{(1)} \right) \]

\[ h_2 = \text{ReLU} \left( w_1^{(1)} x_1 + w_2^{(1)} x_2 + b_2^{(1)} \right) \]

Assume, for a particular training token, that \( h_1 > 0 \) and \( h_2 > 0 \). For that particular training token, what is \( \frac{df}{dw_1^{(1)}} \)? Express your answer in terms of \( x_j, h_j, w_{j,k}^{(l)}, \) and/or \( b_k^{(l)} \) for any values of \( j, k, \) and/or \( l \) that may be useful to you.
Question 7  (7 points)
In the real world, the \((x, y, z)\) coordinates of Joe’s face and Mike’s face are \((14, 3, 7)\) and \((14, 3, 17)\), respectively, where \(z\) is distance from the camera. In the image plane \((x', y')\), which person (Joe or Mike) is closer to the center of the image (the point \((x', y') = (0, 0)\)), and why?