UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

CS440/ECE448 Artificial Intelligence

Conflict Exam 1

Spring 2022

Spring 2022

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

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

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

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

$$\textbf{Perceptron: } \vec{w}_y = \vec{w}_y + \eta\vec{x}, \;\; \vec{w}_{f(\vec{x})} = \vec{w}_{f(\vec{x})} - \eta\vec{x}$$

$$\textbf{Linear Regression w/SGD: } \vec{w} \leftarrow \vec{w} - \frac{\eta}{2}\nabla_{\vec{w}}\varepsilon_i^2 = \vec{w} - \eta\varepsilon_i\vec{x}_i$$

$$\textbf{Logistic Regression: } \nabla_{\vec{w}_c}\mathscr{L}_i = \nabla_{\vec{w}_c}\left(-\ln\frac{e^{\vec{w}_{c_i}^T\vec{x}_i}}{\sum_k e^{\vec{w}_k^T\vec{x}_i}}\right) = \left(\frac{e^{\vec{w}_{c_i}^T\vec{x}_i}}{\sum_k e^{\vec{w}_k^T\vec{x}_i}} - y_{i,c}\right)\vec{x}_i$$

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

$$\textbf{Back-Propagation: } \frac{\partial\mathscr{L}}{\partial h_k^{(l-1)}} = \sum_j \frac{\partial\mathscr{L}}{\partial h_j^{(l)}}\frac{\partial h_j^{(l)}}{\partial h_k^{(l-1)}}$$

$$\textbf{Pinhole Camera: } \frac{x'}{f} = -\frac{x}{z}, \quad \frac{y'}{f} = -\frac{y}{z}$$

Question 1 ***(7 points)***

Consider two binary random variables, X and Y. Suppose that

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

In terms of b and/or c, what is the largest possible value of $P(X=1)$?

Solution:

$$\begin{aligned} P(X=1) &= P(X=1,Y=0) + P(X=1,Y=1) \\ &\leq P(X=1,Y=0) + P(Y=1) \\ &= P(X=1,Y=0) + (1 - P(Y=0)) \\ &= 1 - b + c \end{aligned}$$

Question 2 ***(7 points)***

Suppose you are training a naïve Bayes model. There are two classes, $Y = 0$ and $Y = 1$, with the following observations:

- Training text for class $Y = 0$: "apple apple apple apple apple apple".
- Training text for class $Y = 1$: "banana banana banana banana banana apple".

Use this example to discuss, in a few sentences, the importance of Laplace smoothing.

Solution: Without Laplace smoothing, using these training data, the probability of the word "banana" given class $Y = 0$ would be zero. A reasonable person might suppose that the sentence "apple apple apple banana" is from class $Y = 0$, but the model would assign it zero probability, because it contains the word "banana." Laplace smoothing is important because it gives a small nonzero probability to words that were never seen during training, so it would be possible to label the sentence "apple apple apple banana" as being from class $Y = 0$.

Question 3 ***(7 points)***

Describe, in one sentence each, (1) what does it mean for a classifier to overfit a training corpus?, (2) what does it mean for a model to underfit a training corpus?, (3) how can overfitting and underfitting be avoided?

Solution: Overfitting is when the model learns details of the training corpus that do not generalize to test data. Underfitting is when the model does not learn enough details about the training corpus. Overfitting and underfitting can be avoided by increasing the number of trainable parameters in the model until it reaches minimum error on an independent development test set.

Question 4 ***(7 points)***

Imagine training a perceptron with a training dataset that contains only two training tokens: $\vec{x}_1 = [1,1]^T, y_1 = 1$ and $\vec{x}_2 = [-1,-1]^T, y_2 = -1$. Suppose you begin with the weight vector $\vec{w} = [0,0]^T$ and bias $b = -1$, then present the data in alternating order $\{(\vec{x}_1, y_1), (\vec{x}_2, y_2), (\vec{x}_1, y_1), (\vec{x}_2, y_2), \ldots\}$, with a learning rate of $\eta = 1$, until $\vec{w}$ and b converge. What are the final converged values of $\vec{w}$ and b?

Solution:

$$\vec{w} = [1,1]^T$$
$$b = 0$$

Question 5 ***(7 points)***

In stochastic gradient descent, we train using one training token at a time. Suppose x is a scalar input, and suppose we have

$$\mathscr{L} = -\ln f_2(x)$$

where

$$f_k(\vec{x}) = \frac{e^{w_k x + b_k}}{e^{w_1 x + b_1} + e^{w_2 x + b_2}} \text{ for } k \in \{1,2\}$$

In terms of x, w_1, w_2, b_1, b_2, $f_1(x)$ and/or $f_2(x)$, what is $\frac{d\mathscr{L}}{db_1}$?

Solution:

$$\frac{d\mathscr{L}}{db_1} = f_1(x) = \frac{e^{w_1 x + b_1}}{e^{w_1 x + b_1} + e^{w_2 x + b_2}}$$

Question 6 ***(7 points)***

Consider a two-layer neural network with a scalar input, x. Assume that all of the weights and biases are nonzero, and that the output $f(x)$ is computed as:

$$f(x) = w_{1,1}^{(2)} h_1 + w_{1,2}^{(2)} h_2 + b^{(2)}$$
$$h_1 = \text{ReLU}\left(w_{1,1}^{(1)} x + b_1^{(1)}\right)$$
$$h_2 = \text{ReLU}\left(w_{2,1}^{(1)} x + b_2^{(1)}\right)$$

For what values of x is $\frac{\partial f}{\partial w_{1,1}^{(1)}} \neq 0$? Express your answer in terms of 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.

Solution:

$$\frac{\partial f}{\partial w_{1,1}^{(1)}} = \frac{\partial f}{\partial h_1} \frac{\partial h_1}{\partial w_{1,1}^{(1)}}$$

The first derivative is $\frac{\partial f}{\partial h_1} = w_{1,1}^{(1)}$, which the problem statement declares to be nonzero. The second derivative is nonzero only if $w_{1,1}^{(1)} x + b_1^{(1)} > 0$, i.e.

$$x > -\frac{b_1^{(1)}}{w_{1,1}^{(1)}}$$

Question 7 ***(7 points)***

You are standing on a downward-sloping hillside, with your camera pointed straight ahead of you. Parallel to your line of sight, on your left-hand side (at position $x = -2$ meters), there is a low fence (height 1 meter). The fence descends the hill in front of you, vanishing into a point far in the distance. Let (x', y') denote the position of the fence's vanishing point on your photograph, where x' is horizontal position, y' is vertical position, and $(0,0)$ is the point directly corresponding to your line of sight.

- Is $x' < 0$, $x' = 0$, or $x' > 0$? Explain.
- Is $y' < 0$, $y' = 0$, or $y' > 0$? Explain.

Solution:

- $x' = 0$. The fence is parallel to your line of sight, thus $x = -b$ for some offset b. If we divide by z, and let z go to infinity, we find that $x' = 0$.
- $y' > 0$. The fence is descending, so $y = az + c$ for some negative value of a. Dividing by z, substituting $y/z = -y'/f$, and letting z go to infinity, we find that $y' = -af$, which is positive.