UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

CS440/ECE448 Artificial Intelligence

Exam 3

Spring 2023

May 9, 2023

Your Name: ______________________________

Your NetID: ______________________________

Instructions

- Please write your name on the top of every page.
- Have your ID ready; you will need to show it when you turn in your exam.
- This will be a CLOSED BOOK, CLOSED NOTES exam. You are permitted to bring and use only one 8.5x11 page of notes, front and back, handwritten or typed in a font size comparable to handwriting.
- No electronic devices (phones, tablets, calculators, computers etc.) are allowed.
- SHOW YOUR WORK. Correct answers derivation may not receive full credit if you don't show your work.
- Make sure that your answer includes only the variables that it should include, but DO NOT simplify explicit numerical expressions. For example, the answer $x = \frac{1}{1+\exp(-0.1)}$ is MUCH preferred (much easier for us to grade) than the answer $x = 0.524979$.

Possibly Useful Formulas

$$P(X=x|Y=y)P(Y=y)=P(Y=y|X=x)P(X=x)$$
$$P(X=x)=\sum_y P(X=x,Y=y)$$
$$E\left[f(X,Y)\right]=\sum_{x,y} f(x,y)P(X=x,Y=y)$$

Precision,Recall $= \dfrac{TP}{TP+FP}, \dfrac{TP}{TP+FN}$

MPE=MAP: $f(x)=\arg\max\left(\log P(Y=y)+\log P(X=x|Y=y)\right)$

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

Laplace Smoothing: $P(X=x|Y=y)=\dfrac{\text{Count}(X=x,Y=y)+k}{\text{Count}(Y=y)+k|X|}$, $|X|=\#$ possible distinct values of X

Fairness: $P(Y|A)=\dfrac{P(Y|\hat{Y},A)P(\hat{Y}|A)}{P(\hat{Y}|Y,A)}$

Linear Regression: $\varepsilon_i = f(x_i)-y_i = b+w@x_i-y_i$

Mean Squared Error: $\text{MSE}=\dfrac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2$

Linear Classifier: $f(x)=\arg\max_k w_k@x+b$

Cross-Entropy: $\mathscr{L}=-\dfrac{1}{n}\sum_{i=1}^{n}\log f_{y_i}(x_i)$

Softmax: $\operatorname{soft}\max_c(w@x+b)=\dfrac{\exp(w_c@x+b_c)}{\sum_{k=0}^{V-1}\exp(w_k@x+b_k)}$

Softmax Error: $\varepsilon_{i,c}=\begin{cases} f_c(x_i)-1 & c=y_i \\ f_c(x_i)-0 & \text{otherwise}\end{cases}$

Gradient Descent: $w\leftarrow w-\eta\nabla_w\mathscr{L}$

Neural Net: $h=\text{ReLU}\left(b_0+w_0@x\right)$, $f=\operatorname{soft}\max\left(b_1+w_1@h\right)$

Back-Propagation: $\dfrac{\partial\mathscr{L}}{\partial h_j}=\sum_k\dfrac{\partial\mathscr{L}}{\partial f_k}\times\dfrac{\partial f_k}{\partial h_j}$, $\dfrac{\partial\mathscr{L}}{\partial w_{0,k,j}}=\dfrac{\partial\mathscr{L}}{\partial h_k}\times\dfrac{\partial h_k}{\partial w_{0,k,j}}$

Consistent Heuristic: $h(p)\le d(p,r)+h(r)$

Alpha-Beta Max Node: $v=\max(v,\text{child})$; $\alpha=\max(\alpha,\text{child})$

Alpha-Beta Min Node: $v=\min(v,\text{child})$; $\beta=\min(\beta,\text{child})$

Variance Network: $\mathscr{L}=\dfrac{1}{n-1}\sum_{i=1}^{n}\left(f_2(x_i)-(f_1(x_i)-x_i)^2\right)^2$

Unification: $U = S(P) = S(Q);\ U \Rightarrow \exists x : Q;\ U \Rightarrow \exists x : P$

Bayes Rule: $P(Y=y|X=x) = \dfrac{P(X=x|Y=y)P(Y=y)}{\sum_{y'} P(X=x|Y=y')P(Y=y')}$

Unnormalized Relevance: $\tilde{R}(f_c, x_d) = \dfrac{\partial f_c}{\partial x_d} x_d f_c$

Normalized Relevance: $R(f_c, x_d) = \dfrac{\frac{\partial f_c}{\partial x_d} x_d}{\sum_{d'} \frac{\partial f_c}{\partial x_{d'}} x_{d'}} f_c$

Softmax: $\underset{j}{\text{softmax}}(e) = \dfrac{\exp(e_j)}{\sum_k \exp(e_k)}$

Softmax Deriv: $\dfrac{\partial\, \text{softmax}_m(e)}{\partial e_n} = \underset{m}{\text{softmax}}(e)\delta[m-n] - \underset{m}{\text{softmax}}(e)\,\underset{n}{\text{softmax}}(e),\quad \delta[m-n] = \begin{cases} 1 & m = n \\ 0 & m \neq n \end{cases}$

Viterbi: $v_t(j) = \max_i v_{t-1}(i) a_{i,j} b_j(x_t)$

Transformer: $c_i = \text{softmax}(q_i @ k^T) @ v$

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

Convolution: $w_{k,l} * x_{k,l} = \sum_i \sum_j w_{k-i,l-j} x_{i,j}$

Kalman Prediction : $\mu_{t|t-1} = \mu_{t-1|t-1} + \mu_\Delta,\quad \sigma^2_{t|t-1} = \sigma^2_{t-1|t-1} + \sigma^2_\Delta$

Kalman Gain : $k_t = \dfrac{\sigma^2_{t|t-1}}{\sigma^2_{t|t-1} + \sigma^2_\varepsilon},\quad \sigma^2_{t|t} = \sigma^2_{t|t-1}(1-k_t)$

Kalman Update: $\mu_{t|t} = \mu_{t|t-1} + k_t\left(x_t - \left(\mu_{t|t-1} + \mu_\varepsilon\right)\right)$

Bellman Equation: $U(s) = R(s) + \gamma \max_a \sum_{s'} P(s'|s,a) U(s')$

Value Iteration: $U_i(s) = R(s) + \gamma \max_a \sum_{s'} P(s'|s,a) U_{i-1}(s')$

Policy Evaluation: $U_i(s) = R(s) + \gamma \sum_{s'} P(s'|s,\pi_i(s)) U_i(s')$

Policy Improvement: $\pi_{i+1}(s) = \arg\max_a R(s) + \gamma \sum_{s'} P(s'|s,a) U_i(s')$

Q-Learning: $Q_{t+1}(s,a) = Q_t(s,a) + \alpha\left(Q_{\text{local}}(s,a) - Q_t(s,a)\right)$

TD Learning: $Q_{\text{local}}(s_t,a_t) = R(s_t) + \gamma \max_{a'} Q_t(s_{t+1}, a')$

SARSA: $Q_{\text{local}}(s_t,a_t) = R(s_t) + \gamma Q_t(s_{t+1}, a_{t+1})$

Imitation Learning: $\mathscr{L} = -\log \pi_a(s)$

Deep Q Learning: $\mathscr{L} = \dfrac{1}{2}\left(Q_t(s,a) - Q_{\text{local}}(s,a)\right)^2$

Actor-Critic: $\mathscr{L} = -\sum_a \pi_a(s) Q(s,a)$

Inverse Kinematics: $\mathscr{C}_{\text{obs}} = \{q : \exists b : \phi_b(q) \in \mathscr{W}_{\text{obs}}\}$