

Formula Sheet, CS 440/ECE 448 Exam 1

Probability

$$P(X = x) = \Pr(X = x) \quad \dots \quad \text{or} \quad \dots \quad P(X = x) = \frac{d}{dx} \Pr(X \leq x)$$

$$P(X, Y) = P(X|Y)P(Y)$$

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

Decision Theory

$$f(x) = \underset{y}{\operatorname{argmax}} P(Y = y|X = x), \quad \text{Bayes Error Rate} = \sum_x P(X = x) \min_y P(Y \neq y|X = x)$$

$$\text{Precision} = P(Y = 1|f(X) = 1) = \frac{TP}{TP + FP}$$

$$\text{Recall} = \text{Sensitivity} = P(f(X) = 1|Y = 1) = \frac{TP}{TP + FN}$$

$$\text{Selectivity} = P(f(X) = 0|Y = 0) = \frac{TN}{TN + FP}$$

Naïve Bayes

$$f(x) \approx \underset{y}{\operatorname{argmax}} \left(\log P(Y = y) + \sum_{i=1}^n \log P(W = w_i|Y = y) \right)$$

$$P(W = w_i|Y = y) = \frac{k + \text{Count}(w_i, y)}{k + \sum_{v \in V} (k + \text{Count}(v, y))}$$

Hidden Markov Model

$$v_1(j) = \pi(j)b_j(\mathbf{x}_1)$$

$$v_t(j) = \max_i v_{t-1}(i) a_{i,j} b_j(\mathbf{x}_t), \quad \psi_t(j) = \underset{i}{\operatorname{argmax}} v_{t-1}(i) a_{i,j} b_j(\mathbf{x}_t)$$

$$y^*(T) = \underset{i}{\operatorname{argmax}} v_T(i), \quad y^*(t) = \psi_{t+1}(y^*(t+1))$$

Fairness

- Demographic Parity: $P(f(X)|A = 1) = P(f(X)|A = 0)$
- Equal Odds: $P(f(X)|Y, A = 1) = P(f(X)|Y, A = 0)$
- Predictive Parity: $P(Y|f(X), A = 1) = P(Y|f(X), A = 0)$

Learning

$$\mathcal{R} = E[\ell(Y, f(X))]$$

$$\mathcal{R}_{\text{emp}} = \frac{1}{n} \sum_{i=1}^n \ell(y_i, f(x_i))$$

Linear Regression

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$$

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^n \mathcal{L}_i, \quad \mathcal{L}_i = \frac{1}{2} \epsilon_i^2, \quad \epsilon_i = f(\mathbf{x}_i) - y_i$$

$$\mathbf{w} \leftarrow \mathbf{w} - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{w}}, \quad \frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \frac{1}{n} \sum_{i=1}^n \epsilon_i \mathbf{x}_i$$

Softmax & Sigmoid

$$f_c(\mathbf{x}) = \frac{\exp(\mathbf{w}_c^T \mathbf{x})}{\sum_{k=1}^v \exp(\mathbf{w}_k^T \mathbf{x})} \approx \Pr(Y = c | \mathbf{x})$$

$$\sigma(\mathbf{w}^T \mathbf{x}) = \frac{1}{1 + e^{-(\mathbf{w}^T \mathbf{x})}} \approx \Pr(Y = 1 | \mathbf{x})$$

$$\mathcal{L} = -\ln f_y(\mathbf{x}), \quad \frac{\partial \mathcal{L}}{\partial f_c(\mathbf{x})} = \begin{cases} -\frac{1}{f_c(\mathbf{x})} & c = y \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{w}_c \leftarrow \mathbf{w}_c - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{w}_c} = \mathbf{w}_c - \eta \epsilon_c \mathbf{x}, \quad \epsilon_c = \begin{cases} f_c(x_i) - 1 & c = y \\ f_c(x_i) - 0 & \text{otherwise} \end{cases}$$

Multi-Layer

$$f_k = \text{softmax}_k \mathbf{z}^{(2)}, \quad \mathbf{z}_k^{(2)} = \sum_{j=1}^n w_{k,j}^{(2)} h_j, \quad h_j = \text{ReLU} \left(\sum_{i=1}^d w_{j,i}^{(1)} x_i \right)$$

$$\frac{\partial \mathcal{L}}{\partial w_{j,i}^{(1)}} = \sum_{k=1}^v \left(\frac{\partial \mathcal{L}}{\partial z_k^{(2)}} \right) \left(\frac{\partial z_k^{(2)}}{\partial h_j} \right) \left(\frac{\partial h_j}{\partial w_{j,i}^{(1)}} \right) = \sum_{k=1}^v (f_k - 1_{y=k}) w_{k,j}^{(2)} 1_{h_j > 0} x_i$$