Midterm Practice Problems

CS 446/ECE 449

March 5, 2024

1. Is it possible to use a linear regression model for binary classification? If so, how do we map the regression output $\mathbf{w}^\top \mathbf{x}$ to the class labels $y \in \{-1, 1\}$?

2. Consider a model with the following parameterization:

$$p(y^{(i)}|\mathbf{x}) = \frac{1}{1 + \exp(-\mathbf{w}^\intercal \mathbf{x}^{(i)} - b)}, \tag{1}$$

where $\mathbf{w} \in \mathbb{R}^2$ and $b \in \mathbb{R}$.

What is the highest accuracy for this model on the XOR dataset? **Note:** To compute accuracy, we use a threshold of 0.5, *i.e.*, the final prediction of the model is $\delta[p(y^{(i)}|\mathbf{x}) > 0.5]$, where δ denotes the indicator function.

3. Consider another model with the parametrization shown below:

$$\tilde{y}^{(i)} = \frac{1}{1 + \exp(-a_2^{(i)})} \tag{2}$$

$$a_2^{(i)} = \theta^\intercal \max(\mathbf{a_1^{(i)}}, 0) + b \tag{3}$$

$$\mathbf{a_1^{(i)}} = \mathbf{W}\mathbf{x^{(i)}} + \mathbf{c} \tag{4}$$

where $\theta \in \mathbb{R}^2$, $b \in \mathbb{R}$, $\mathbf{W} \in \mathbb{R}^{2\times 2}$ and $\mathbf{c} \in \mathbb{R}^2$.

Find a θ and b that achieve 100 % accuracy on the XOR dataset, given $\mathbf{W} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $\mathbf{c} = [0, -1]^\intercal$. *To show your work, write out $\mathbf{a_1^{(i)}}$ and $a_2^{(i)}$ for the four datapoints in the XOR dataset and your choice of θ, and b.*

Note: To compute accuracy, we use a threshold of 0.5, *i.e.*, the final prediction of the model is $\delta[\tilde{y}^{(i)} > 0.5]$, where δ denotes the indicator function.

More questions, variations on HWs.

1. HW1: Q4.2 Derive the same, but now the features have covariance $\Sigma \neq I$ where Σ is diagonal. What if Σ were not diagonal?

2. HW2: Q3.5 What if the decision stumps were instead "diagonal", i.e. the split rule is $\boldsymbol{x} = (x_1, x_2) \geq (\tau, \tau)$ instead of just an axis-aligned splits?

3. HW2: Q4.1 Say that our hypothesis space $\mathcal{H}$ contains 4 functions: h_1 always guesses heads. h_2 guesses heads 1/2 of the time. h_3 guesses heads 1/4 of the time, and h_4 always guesses tails. How many samples are needed so that with confidence 95%, none of the $h \in \mathcal{H}$ have an accuracy of $|R(h) - \hat{R}_S(h)| > 0.05$?

4. HW3: Q1.4 What are the dimensions of $\frac{\partial \ell(f)}{\partial z}$ for $z = \sigma(w_0^\top x)$? What about $\frac{\partial \ell(f)}{\partial w_0}$ and $\frac{\partial \ell(f)}{\partial w_1}$?