Due Date: 1400.02.03

Homework4

Theoretical

1. (10 points) Suppose that we have a linearly separable dataset. Show that non-batch perceptron algorithm will converge in finite steps. (Let the initial value of w be **0**)

2. (15 points) Consider using a logistic regression model $h_\theta(x) = g(\theta^T x)$ where g is sigmoid function, and let a training set $\{(x^{(i)}, y^{(i)}; i = 1, \cdots, m\}$ be given as usual. The maximum likelihood estimate of the parameters θ is given by

$$\theta_{\text{ML}} = \arg\max_\theta \prod_{i=1}^{m} p(y^{(i)}|x^{(i)}; \theta)$$

If we wanted to regularize logistic regression, then we might put a Bayesian prior on the parameters. Suppose we chose the prior $\theta \sim \mathcal{N}(0, \tau^2 I)$ (here, $\tau > 0$, and I is the $n+1$-by-$n+1$ identity matrix), and then found the MAP estimate of θ as:

$$\theta_{\text{MAP}} = \arg\max_\theta p(\theta) \prod_{i=1}^{m} p(y^{(i)}|x^{(i)}, \theta)$$

Prove that

$$\|\theta_{\text{MAP}}\|_2 \leqslant \|\theta_{\text{ML}}\|_2$$

3. (12 points) If we have a binary target value ($\forall i, y_i \in \{0, 1\}$) then show that Naive Bayes classifier is a linear classifier.

4. (8 points) Suppose that the true regression function that we are trying to predict is $h(x)$ and we have M predictive models $y_1(x), \cdots, y_M(x)$. If $H_M(x)$ is given by

$$H_M(x) = \frac{1}{M} \sum_{i=1}^{M} y_i(x)$$

then prove that

$$\mathbb{E}_x[(H_M(x) - h(x))^2] \leqslant \frac{1}{M} \sum_{i=1}^{M} \mathbb{E}_x[(h_i(x) - h(x))^2]$$

5. (15 points) We want to train a decision tree on a dataset with binary features using ID3 algorithm.

- **(a)** Prove that training error is 0.
- **(b)** If we have M features, is part (a) still true for trees with depth $M - 1$?

Practical

You are going to use Blood Transfusion Service Center dataset[1] in this assignment. You are not allowed to use sklearn. Please hand in a report and your code (You can also use Jupyter Notebook instead).

1. (20 points) Train a logistic regression classifier and a decision tree to classify the dataset. Split the dataset to train and test data to check if any of the models are overfitting. Report accuracy, recall and F1-score for both of the models.

2. (10 points) As you see in part 1, F1-score and recall are very low. What is the reason? How can you solve this problem?

3. (10 points) Use your answers in part 2 and change the models in a way that we have higher F1-score.

4. (15 points) Train an AdaBoost classifier that each weak learner is a stump (Stumps are decision trees with depth one). Report accuracy, recall and F1-score. If F1-score is low, try to improve it using your answer for part 2.

[1]For more information about the dataset, use this link.