Assignment 4 Solutions

ML Class: CS 6375.501

November 23, 2020

1 Assignment Policies for CS 6375

The following are the policies regarding this assignment.

1. This assignment needs be done individually by everyone.
2. You are expected to work on the assignments on your own. If I find the assignments of a group of (two or more) students very similar, the group will get zero points towards this assignment. You may possibly also be reported to the judiciary committee.
3. Please use Python for writing code. You can submit the code as a Jupyter notebook
4. For the theory questions, please use Latex
5. This Assignment is for 20 points.
6. This will be due on November 30th.

2 Questions

1. **Variance/Bias Tradeoff (5 points):** Below, you are provided two classifiers, and you need to identify the tradeoff between variance and bias in each case (i.e. for e.g. compare the second classifier to the first and ideltify if the bias/variance is lower or higher)
   - Logistic Regression vs Neural Network
   - Logistic Regression vs Decision Tree
   - Decision Tree vs Random Forest
   - Decision Tree vs Gradient Boosted Tree
   - Logistic Regression vs 1NN classifier
2. **Bagging with Linear Regression (3 Points):** Imagine that instead of Random Forest, which performs bagging on decision trees, we perform bagging on a linear regression model. What will be the algorithm? Will this identical to running a single linear regression? Also, comment on the bias and variance of the bagged linear regression model in comparison with a simple linear regression.
3. **Gradient Boosting (7 points):** In class, we studied gradient boosted decision trees with the squared L2 Loss. Provide the algorithm for gradient boosting if instead of the squared L2 loss, we use the Logistic Loss and the Hinge Loss.

4. **Neural Networks with Logistic Loss and ReLU non-linearities (5 points):** In class, we derived the back-propogation and gradient descent expressions for Squared L2 Loss (linear regression loss) with sigmoid activation. Derive the reccursive expressions if instead, we have the logistic loss (assume binary classification) and ReLU non-linearities.

Solutions

Q1 a) Logistic Regression: (High bias / Low Var)
Neural Network (Low bias / High Var)

b) Logistic Regression v/s DT
high bias — Low bias
Low var — High var

c) DT v/s RF
← same bias →
High var — Low var

d) DT v/s GBDT
Low bias — Lower bias
High var — Low var.

e) LR v/s NN
high bias — low bias
low var — high var.

Q2) $\{(x_1, y_1), \ldots (x_N, y_N)\}$

Lin Reg: $\arg\min_{w,b} \sum_{i=1}^{N} (y_i - w^T x_i - b)^2 = \hat{w}, \hat{b}$

$$\hat{y} = \hat{w} x + \hat{b}$$

Bagging:

for $i = 1 : T$

select random feature set & random instances

$\{(x_j^i, y_j^i)\}_{j=1:N}$

$$\hat{w}_i, \hat{b}_i = \underset{w,b}{\arg\min} \sum_{i=1}^{N} (y_j^i - w^T x_j^i - b)^2$$

$$\hat{y} = \sum_{i=1}^{T} \hat{w}_i x + \hat{b} = \hat{\hat{w}} x + \hat{\hat{b}}$$

where

$$\hat{\omega} = \sum_{i=1}^{T} \hat{w}_i \quad , \quad \hat{b} = \sum_{i=1}^{T} \hat{b}_i$$

$\therefore$ Bagging with Lin-Reg is equivalent to a single Linear Regression.

$\therefore$ Bias & Variance are exactly the same!!

Q3] with Logistic loss

$$L(y, F(x)) = \log(1 + \exp(-y_i F(x_i)))$$

$$\frac{\partial L}{\partial F} = \frac{-y_i \exp(-y_i F(x_i))}{1 + \exp(-y_i F(x_i))}$$

with Hinge loss,

$$L(y, F(x)) = \max(0, 1 - yF(x))$$

Compute $\frac{\partial L}{\partial F} = -y_i \mathbb{1}_{1-yF(x) \geq 0}$

Implement GBDT from slide 31 from lecture 17 in slide notes (see marked version)

Q4)

$$L(F(W_l^T F(W_{l-1}^T F(\ldots)), y))$$

Consider $l = 2$

$$\Rightarrow L(F(W_2^T F(W_1^T x + b_1) + b_2), y)$$

Let $z_2 = W_2^T F(z_1) + b_2$

$$\frac{\partial L}{\partial z_2} = \frac{\partial L(F(z_2), y)}{\partial z_2}$$

Now, $F(z_2) = \max(0, z_2)$

$$L(F(z_2), y) = \log(1 + \exp(-y F(z)))$$

$\therefore \dfrac{\partial L}{\partial z_2} = \dfrac{-y \exp(-y F(z_2))}{1+\exp(-y F(z_2))} \mathbb{1}_{F(z_2) \geq 0}$

substitute w_2, b_2 & set $\dfrac{\partial L}{\partial w_2}, \dfrac{\partial L}{\partial b_2}$.

$\dfrac{\partial z_2}{\partial w_2} = F(z_1), \quad \dfrac{\partial z_2}{\partial b_2} = 1.$

By chain rule,

$$\frac{\partial L}{\partial z_1} = \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial z_1}$$

$$z_2 = w_2^T f(z_1) + b_2$$

$$\frac{\partial z_2}{\partial z_1} = w_2^T \mathbb{1}_{z_1 \geq 0}$$

Finally compute $\dfrac{\partial z_2}{\partial b_1}, \dfrac{\partial z_2}{\partial w_1}$