No.

Date: / /

Quiz 1.2

1. When the data is unbalanced (1% positive & 99% negative) if we use the accuracy to evaluate a model, if the model always predict the mode of label (negative), and we could get 99.9% of accuracy, but the model is not so good. So we F_1 score $= 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$.
where the recall in previous case would be 0%.

2. When we choose activation function we should find a function without vanishing gradient problem and it should be differentiable and the binary classification function is not differentiable.

3. Bias and variance tradeoff is a big issue in machine. High bias can cause the model to neglect the special case and more generalization (but prone to underfitting) High variance would make model focus on special case and reduce the training error (but prone to overfitting)

4. Because we used the skill of "boostraping" to train individual tree, although we could get many random trees, it's not so correlated to each other, so we don't need to prune the tree (to prevent overfitting).

5. When we use sklearn to build a model, some columns might contain catagorical data. assume we have a column =

	Color
1	red
2	blue
3	green

the one hot encoding function (pd.get_dummies)

$\xrightarrow{\text{pd.getdummies}}$

	Color_red	Color_blue	Color_green
1	1	0	0
2	0	1	0
3	0	0	1

(transfer catagorical data to numeric (0, 1))

No.
Date: / /

6, we could use drop out to prevent neural network overfitting (often) in the last layer in a neural we could randomly select some neural's weight to be 0, and it could prevent overly rely in some special case.

Quiz 1.

$$\frac{\partial CE}{\partial \theta} = \frac{\partial CE}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \theta}$$

$$\frac{\partial CE}{\partial W^{(2)}} = \frac{\partial CE}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \theta} \frac{\partial \theta}{\partial W^{(2)}}$$