Quiz2

1 True or False [60 pts]

- Machine learning is the process of discovering patterns in data through runing learning algorithm on hypothesis set. T
- Supervised learning is a type of machine learning where the training data includes the desired outputs. T
- Unsupervised learning is a type of machine learning where the training data does not include the desired outputs. T
- Classification, and regression are all types of supervised learning. T
- For known target functions, we still need to construct a hypothesis set and a learning algorithm to solve the problem. F
- Generally, fewer training samples are better, as they allow any hypothesis set to fit easily. F
- We can use the same learning algorithm to solve for any hypothesis set. F
- A training set is a set data used to discover potentially predictive relationships, so that the larger training set, the better. T
- A test set is a set of data used to assess the strength and utility of a predictive relationship T
- The Expected Risk Minimization refer to minimize the loss on a particular observed sample data. F
- The Empirical Risk Minimization refer to minimize the loss on the true joint distribution. F
- You will work hard and pass the Machine Learning Class T/F

2 ERM [40 pts]

1. Given the dataset $\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$, hypothesis set $\mathcal{H}$, please write down the formula of Empirical Risk Minimization and Expected Risk Minimization.

See Lecture+2-ERM Slides 23.

2. Assume $h_\theta \in \mathcal{H} = \{\theta + 2024;\ \theta \in \mathcal{R}\}$ and if the loss function is defined as loss(h_θ, y_i) $= (h_\theta - y_i)^2$, write down the θ that minimizes the empirical risk.

$\theta = \frac{1}{n}\sum_{i=1}^{n} y_i - 2024$

3. Assume $h_\theta \in \mathcal{H} = \{2024 \times \theta;\ \theta \in \mathcal{R}\}$ and if the loss function is defined as loss(h_θ, y_i) $= |h_\theta - y_i|$, write down the θ that minimizes the empirical risk.

$\theta =$ median y_i / 2024

4. Assume $h_\theta \in \mathcal{H} = \{\theta;\ \theta \in \mathcal{R}\}$ and if the loss function is defined as loss(h_θ, y_i) $= |h_\theta - y_i|_\infty$ (note that $||_\infty$ denotes Infinity (L∞) Norm), write down the θ that minimizes the empirical risk.

$\theta = (\mathrm{y}_{\mathrm{max}} + \mathrm{y}_{\mathrm{min}})$ / 2