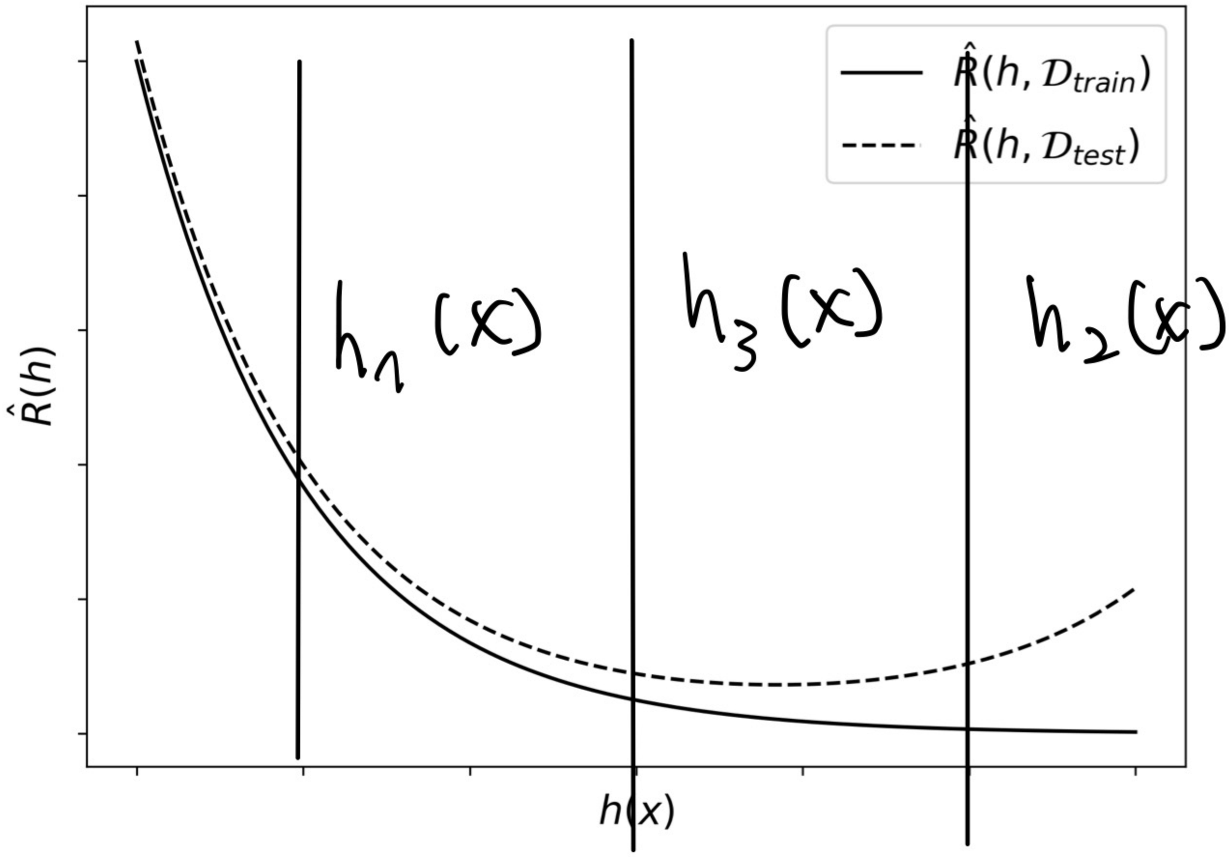
Homework 7

Problem 1

From the given plots of classifiers we can make the followiing statements:

1. h_1 uses a linear boundary to make decision, the model is too simple for the distribution of data set, thus we are in the "under-fitting region" and we have bias problem, therefore the $\hat{R}(h, D_{\text{train}})$ and $\hat{R}(h, D_{\text{test}})$ should be both relatively large;
2. by using high order polynomial we can obtain a classifier like h_2, where overfitting occurs, we have variance problem, thus the $\hat{R}(h, D_{\text{train}})$ shall be small but $\hat{R}(h, D_{\text{test}})$ would be large;
3. h_3 is derived by low order polynomial, it can good represent the distribution of data set, so the $\hat{R}(h, D_{\text{train}})$ and $\hat{R}(h, D_{\text{test}})$ should be both small.

Consider all the arguments above, we can answer the given question as interpreted in the following figure:



Problem 2

First of all, $R(\hat{h})$ should be decreasing as n getting larger. With $\inf_{h\in\mathcal{H}} R(h)$ stands for the minimum possible risk that can be achieved by $h \in \mathcal{H}$, the estimation error should getting smaller w.r.t. the increase of n.

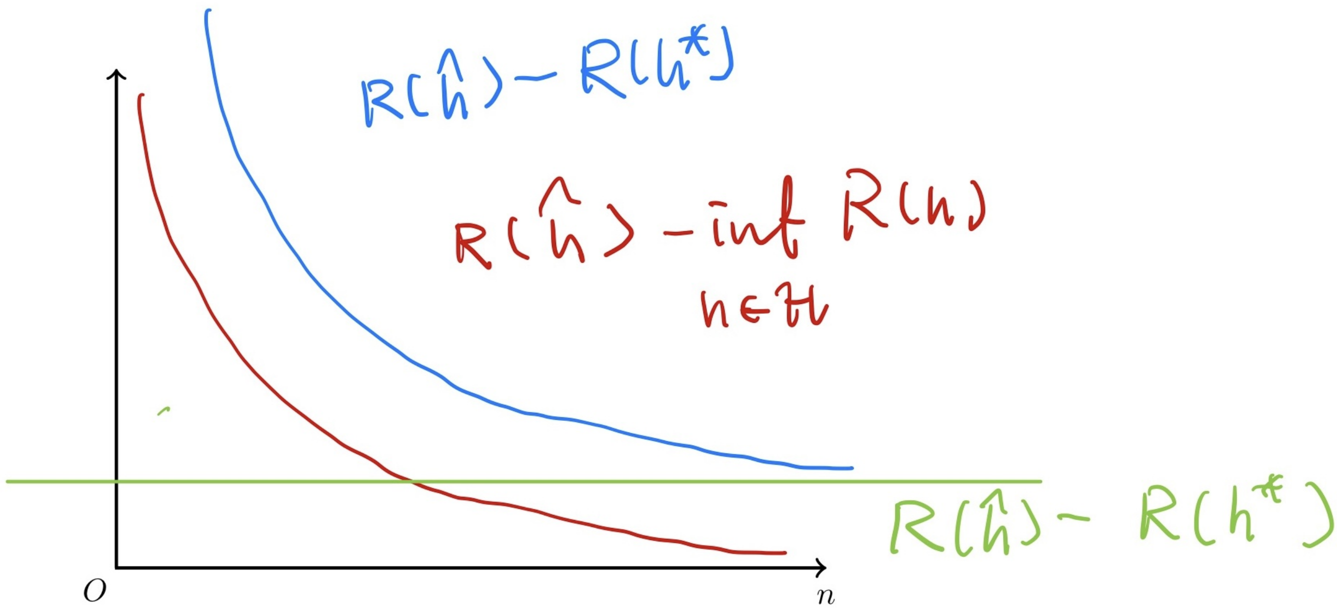
Subsequently, $\inf_{h\in\mathcal{H}} R(h)$ and $R(h^*)$ are both independent with n, so in the plot the approximation error should be constant.

Finally, we have the following equation for the excess error

$$R(\hat{h}) - R(h^*) = (R(\hat{h}) - \inf_{h\in\mathcal{H}} R(h)) + (\inf_{h\in\mathcal{H}} R(h) - R(h^*)),$$

with the term $\inf_{h\in\mathcal{H}} R(h) - R(h^*)$ being a constant, the excess error is just the estimation error shifted by approximation error.

Therefore, we get the following plot



Problem 3

In the lecture note, we have the following equation

$$R(\theta) \le \hat{R}(\theta) + B\sqrt{\frac{\log(\frac{|\mathcal{H}|}{\delta})}{2\cdot n}}$$

Given $\theta \in \mathbb{R}^p$ and each entry of it is either 1 or -1 we have $|\mathcal{H}| = 2^p$, and since we are using 0-1 loss, so the loss is bounded by 1, which means $B = 1$, therefore we have

$$R(\theta) \le \hat{R}(\theta) + \sqrt{\frac{\log(\frac{2^p}{\delta})}{2\cdot n}}$$