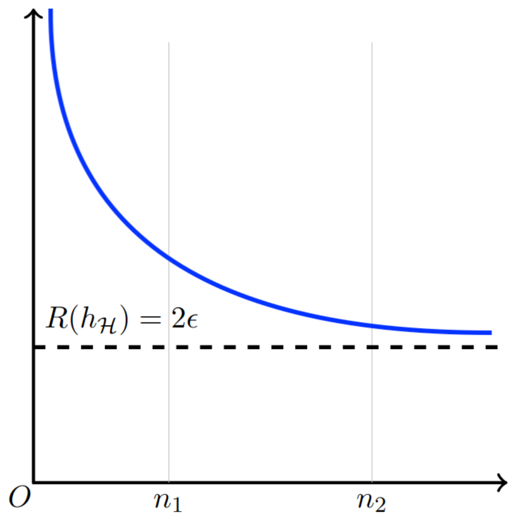
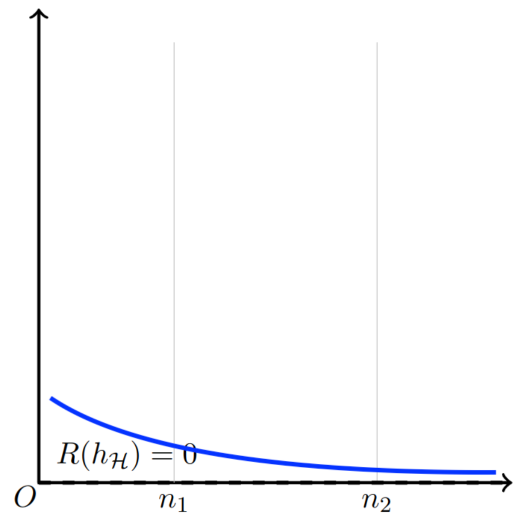
# Homework 9

## Problem 1

The corresponding figures are shown below




(b)




(c)

## Problem 2

### 1

In this setup we use hinge loss as loss function, which is a (non-strict) convex function. Additionally, the empirical risk is a linear transform of the sum of a bunch of convex functions, therefore the given empirical risk is also convex.

### 2

The gradient depends on the value of $y\langle\theta, \mathbf{x}\rangle$:

1. If $y\langle\theta, \mathbf{x}\rangle \geq 1$, then the gradient $\mathbf{g}(\theta, \mathbf{x}_i, y_i) = 0$.
2. If $y\langle\theta, \mathbf{x}\rangle < 1$, then its gradient w.r.t. is $\mathbf{g}(\theta, \mathbf{x}_i, y_i) = y_i\mathbf{x}_i$

The expectation of $\mathbf{g}(\theta, \mathbf{x}_i, y_i) = \nabla_\theta l(\theta, \mathbf{x}_i, y_i)$ is given by

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{g}(\theta, \mathbf{x}_i, y_i)\right] &= \frac{1}{n}\sum_{i=1}^{n} \nabla_\theta l(\theta, \mathbf{x}_i, y_i) \\
&= \sum_{i=1}^{n} \frac{1}{n} \nabla_\theta l(\theta, \mathbf{x}_i, y_i) \\
&= \nabla_\theta \hat{R}(\theta)
\end{aligned}
$$

# 3

The given condition can be interpreted as

$$
\begin{aligned}
P\left[\hat{R}(\theta^n) \leq \epsilon_1\right] \geq 1 - \delta_1 \\
P\left[\hat{R}(\theta^n) \geq \epsilon_1\right] \leq \delta_1
\end{aligned}
$$

and

$$
\begin{aligned}
P\left[\sup_\theta \left(R(\theta) - \hat{R}(\theta)\right) \leq \epsilon_2\right] \geq 1 - \delta_2 \\
P\left[\sup_\theta \left(R(\theta) - \hat{R}(\theta)\right) \geq \epsilon_2\right] \leq \delta_2
\end{aligned}
$$

The goal is to prove

$$
P\left[R(\theta^n) \leq \epsilon_1 + \epsilon_2\right] \geq 1 - (\delta_1 + \delta_2),
$$

this can be further converted into the proof of

$$
P\left[R(\theta^n) \geq \epsilon_1 + \epsilon_2\right] \leq (\delta_1 + \delta_2).
$$

and the left hand side of the inequalty can be interpreted as

$$
\begin{aligned}
P\left[R(\theta^n) \geq \epsilon_1 + \epsilon_2\right] &= P\left[\hat{R}(\theta^n) \geq \epsilon_1 \ \cup \ \sup_\theta \left(R(\theta) - \hat{R}(\theta)\right) \geq \epsilon_2\right] \\
&= P\left[\hat{R}(\theta^n) \geq \epsilon_1\right] + P\left[\sup_\theta \left(R(\theta) - \hat{R}(\theta)\right) \geq \epsilon_2\right] \\
&\leq \delta_1 + \delta_2.
\end{aligned}
$$

Thus, the given equation is prooven.