Machine Learning Theory
Tübingen University, WS 2016/2017
Assignment 4

Due: Monday, January 16th, 12:15

Hand in at beginning of tutorial!

The solutions to this assignment will be discussed in the tutorial on Monday, Jan 16. If you want your assignment to be marked, it has to be handed in at the beginning of the tutorial. If you can not make it to the tutorial yourself, please arrange for one of your classmates to hand in your assignment. In this assignment you can get 25 points. You can gain up to 20 bonus points by solving the last two problems.

To better understand the questions, refer to Lecture 9. To gain full marks, explain your answers carefully and prove your claims.

1. **Rademacher complexity is nonnegative** Prove that the conditional Rademacher complexity is nonnegative, i.e.

$$\hat{R}_n(\mathcal{F}) = \mathbb{E}_{\sigma_1,\dots,\sigma_n}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n \sigma_i f(Z_i)\middle| Z_1,\dots,Z_n\right] \geq 0$$

for any function class $\mathcal{F}$ defined over any input space $\mathcal{Z}$ and any points $Z_1,\dots,Z_n \in \mathcal{Z}$.

(5 marks)

2. **Rademacher complexity is monotonic** Take two classes $\mathcal{F}$ and $\mathcal{F}'$ of functions, defined over $\mathcal{Z}$ and mapping to $\mathbb{R}$. Prove that

$$\max\big(\hat{R}_n(\mathcal{F}), \hat{R}_n(\mathcal{F}')\big) \leq \hat{R}_n(\mathcal{F}\cup\mathcal{F}').$$

Moreover, define the following version of RC, including an absolute value:

$$\hat{R}^a_n(\mathcal{F}) = \mathbb{E}_{\sigma_1,\dots,\sigma_n}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\left|\sum_{i=1}^n \sigma_i f(Z_i)\right| \middle| Z_1,\dots,Z_n\right] \geq 0.$$

Show that

$$\hat{R}_n(\mathcal{F}) \leq \hat{R}^a_n(\mathcal{F}).$$

Finally, prove that

$$\hat{R}^a_n(\mathcal{F}\cup\mathcal{F}') \leq \hat{R}^a_n(\mathcal{F}) + \hat{R}^a_n(\mathcal{F}')$$

while this inequality does not generally hold for $\hat{R}_n(\mathcal{F})$. **(10 marks)**

3. **McDiarimd's inequality generalizes Hoeffding's inequality** Prove that McDiarmid's inequality for functions of bounded differences (refer to Lecture 9) implies Hoeffding's inequality for sums of random variables (refer to Lecture 3).

(5 marks)

4. **Uniform deviations have bounded differences** Take any input space $\mathcal{X}$ and output space $\mathcal{Y}$. Consider any distribution P on $\mathcal{X} \times \mathcal{Y}$. Denote $Z_i = (X_i, Y_i)$, where $(X_1, Y_1), \ldots, (X_n, Y_n) \in \mathcal{X} \times \mathcal{Y}$ are sampled i.i.d. from P. Take any class of predictors $\mathcal{H}$ mapping $\mathcal{X}$ to $\mathcal{Y}$ and any loss function $\ell(y, y')$ such that $\ell\big(h(X), Y\big) \in [0, M]$ for any $h \in \mathcal{H}$, $X \in \mathcal{X}$, and $Y \in \mathcal{Y}$.

$$\begin{aligned} f(Z_1, \ldots, Z_n) &= \sup_{h \in \mathcal{H}} \big(L(h) - L_n(h)\big) \\ &= \sup_{h \in \mathcal{H}} \left(\mathbb{E}_{(X,Y) \sim P}\big[\ell\big(h(X), Y\big)\big] - \frac{1}{n} \sum_{i=1}^{n} \ell\big(h(X_i), Y_i\big) \right). \end{aligned}$$

Prove that f satisfies the bounded difference condition (see lecture 9, McDiarmid's inequality) with parameters $c_i = M/n$.

(5 marks)

5. **(*) Symmetrization inequality** Prove Theorem 4 of Lecture 9. Read the lecture for a hint.

(10 marks)

6. **(*) Bayes optimal predictors in squared regression** Take any input space $\mathcal{X}$ and assume the output space is $\mathcal{Y} = \mathbb{R}$. Take any distribution P over $\mathcal{X} \times \mathcal{Y}$, such that

$$P(-M \leq Y \leq M) = 1$$

for some constant $M > 0$. Prove that for any function $f \colon \mathcal{X} \to \mathbb{R}$ the following inequality holds:

$$\mathbb{E}_{(X,Y) \sim P}\left[\big(f(X) - Y\big)^2\right] \geq \mathbb{E}_{(X,Y) \sim P}\left[\big(f^*(X) - Y\big)^2,\right]$$

where $f^*(x) = \mathbb{E}[Y|X = x] = \int_{\mathbb{R}} Y dP(Y|X = x)$ is the conditional expectation of Y given $X = x$. This shows that the conditional expectation is a Bayes optimal predictor for the squared loss.

(10 marks)