# Homework #2

RELEASE DATE: 11/14/2017

DUE DATE: 12/12/2017, BEFORE 14:00

QUESTIONS ABOUT HOMEWORK MATERIALS ARE WELCOMED ON THE FACEBOOK FORUM.

*Unless granted by the instructor in advance, you must turn in a printed/written copy of your solutions (without the source code) for all problems.*

*For problems marked with (*), please follow the guidelines on the course website and upload your source code to designated places. You are encouraged to (but not required to) include a README to help the TAs check your source code. Any programming language/platform is allowed.*

*Any form of cheating, lying, or plagiarism will not be tolerated. Students can get zero scores and/or fail the class and/or be kicked out of school and/or receive other punishments for those kinds of misconducts.*

*Discussions on course materials and homework solutions are encouraged. But you should write the final solutions alone and understand them fully. Books, notes, and Internet resources can be consulted, but not copied from.*

*Since everyone needs to write the final solutions alone, there is absolutely no need to lend your homework solutions and/or source codes to your classmates at any time. In order to maximize the level of fairness in this class, lending and borrowing homework solutions are both regarded as dishonest behaviors and will be punished according to the honesty policy.*

*You should write your solutions in English or Chinese with the common math notations introduced in class or in the problems. We do not accept solutions written in any other languages.*

> This homework set comes with 200 points and 20 bonus points. In general, every homework set would come with a full credit of 200 points, with some possible bonus points.

**1.** (60 points) Go register for the Coursera version of the first part of the class ( `https://www.coursera.org/teach/ntumlone-mathematicalfoundations/` ) and solve its homework 2. The registration should be totally free. Then, record the highest score that you get within up to 3 trials. Please print out a snapshot of your score as an evidence. (*Hint: The problems below are simple extensions of the Coursera problems.*)

**2.** (20 points) Derive the growth function $m_{\mathcal{H}}(N)$ of "positive-and-negative intervals on $\mathbb{R}$." The hypothesis set $\mathcal{H}$ of "positive-and-negative intervals" contains the functions which are $+1$ within an interval $[\ell, r]$ and $-1$ elsewhere, as well as the functions which are $-1$ within an interval $[\ell, r]$ and $+1$ elsewhere. (*Hint: For instance, the hypothesis $h_1(x) = sign(x(x-4))$ is a negative interval with $-1$ within $[0, 4]$ and $+1$ elsewhere, and hence belongs to $\mathcal{H}$. The hypothesis $h_2(x) = sign((x+1)(x)(x-1))$ contains two positive intervals in $[-1, 0]$ and $[1, \infty)$ and hence does not belong to $\mathcal{H}$.*)

**3.** (20 points) Consider the "polynomial discriminant" hypothesis set of degree $D$ on $\mathbb{R}$, which is given by

$$\mathcal{H} = \left\{ h_{\mathbf{c}} \;\middle|\; h_{\mathbf{c}}(x) = \text{sign}\left( \sum_{i=0}^{D} c_i x^i \right) \right\}$$

What is the VC-dimension of such an $\mathcal{H}$? Write down your derivation steps.

**4.** (20 points) Consider the "triangle waves" hypothesis set on $\mathbb{R}$, , which is given by

$$\mathcal{H} = \{h_\alpha \mid \quad h_\alpha(x) = \text{sign}(|\alpha x \bmod 4 - 2| - 1), \alpha \in \mathbb{R}\}$$

Here ($x$ mod 4) is a number $x - 4k$ for some integer $k$ such that $x - 4k \in [0, 4)$. For instance, (11.26 mod 4) is 3.26, and ($-11.26$ mod 4) is 0.74. What the VC-Dimension of such an $\mathcal{H}$? Please prove your answer.

**5.** (20 points) When $\mathcal{H}_1 \subseteq \mathcal{H}_2$, prove that $d_{vc}(\mathcal{H}_1) \leq d_{vc}(\mathcal{H}_2)$. (*Hint: This may partially help you solve Questions 14 and 15 on Coursera.*)

**6.** (20 points) Consider $\mathcal{H}_1$ as the positive-ray hypothesis set (as discussed in class), and $\mathcal{H}_2$ as the negative-ray hypothesis set (which contains the negation of each hypothesis in $\mathcal{H}_1$). We showed that $m_{\mathcal{H}_1}(N) = N + 1$ in class. Write down $m_{\mathcal{H}_1 \cup \mathcal{H}_2}(N)$ and use that to calculate $d_{vc}(\mathcal{H}_1 \cup \mathcal{H}_2)$. (*Hint: This may partially help you solve Question 15 on Coursera.*)

**7.** (20 points) For Problems 7–8, you will play with the decision stump algorithm. In class, we taught about the learning model of "positive and negative rays" (which is simply one-dimensional perceptron) for one-dimensional data. The model contains hypotheses of the form:

$$h_{s,\theta}(x) = s \cdot \text{sign}(x - \theta).$$

The model is frequently named the "decision stump" model and is one of the simplest learning models. As shown in class, for one-dimensional data, the VC dimension of the decision stump model is 2.

In fact, the decision stump model is one of the few models that we could easily minimize $E_{in}$ efficiently by enumerating all possible thresholds. In particular, for $N$ examples, there are at most $2N$ dichotomies (see page 22 of lecture 5 slides), and thus at most $2N$ different $E_{in}$ values. We can then easily choose the dichotomy that leads to the lowest $E_{in}$, where ties an be broken by randomly choosing among the lowest $E_{in}$ ones. The chosen dichotomy stands for a combination of some "spot" (range of $\theta$) and $s$, and commonly the median of the range is chosen as the $\theta$ that realizes the dichotomy.

In the next problem, you are asked to implement such and algorithm and run your program on an artificial data set. We shall start by generating a one-dimensional data by the procedure below:

(a) Generate $x$ by a uniform distribution in $[-1, 1]$.

(b) Generate $y$ by $f(x) = \tilde{s}(x) + \text{noise}$ where $\tilde{s}(x) = \text{sign}(x)$ and the noise flips the result with 20% probability.

For any decision stump $h_{s,\theta}$ with $\theta \in [-1, 1]$, express $E_{out}(h_{s,\theta})$ as a function of $\theta$ and $s$. Write down your derivation steps.

**8.** (20 points, *) Generate a data set of size 20 by the procedure above and run the one-dimensional decision stump algorithm on the data set. Record $E_{in}$ and compute $E_{out}$ with the formula above. Repeat the experiment 1000 times and plot a scatter plot of $E_{in}$ versus $E_{out}$. Describe your findings.

# Bonus: More about the Growth Function of Perceptrons

**9.** (Bonus 20 points) For perceptron learning model in $d$ dimensions, prove that the growth function is given by

$$m_{\mathcal{H}}(N) = 2 \sum_{i=0}^{d} \binom{N-1}{i}.$$

*(Warning: this problem is supposed to be very hard.)*