Machine Learning, Spring 2019
Homework 5

Due on 23:59 May 7, 2019

---

1. Submit your solutions to Gradescope (www.gradescope.com). Homework of this week contains two part, **theoretical part** and **programming part**. So there are two assignments in gradescope, the assignment titled with programming part will require you to submit your code while the results of programming part should be put in theoretical part.

2. **Make sure each solution page is assigned to the corresponding problems** when you submit your homework.

3. Any programming language is allowed for your code, but **make sure it is clear and readable with necesary comments.**

---

1 Bias-variance Decomposition

When there is noise in the data, $E_{\text{out}}\left(g^{(\mathcal{D})}\right) = \mathbb{E}_{\mathbf{x},y}\left[\left(g^{(\mathcal{D})}(\mathbf{x}) - y(\mathbf{x})\right)^2\right]$, where $y(\mathbf{x}) = f(\mathbf{x}) + \epsilon$. If ϵ is a zero mean noise random variable with variance σ^2, show that the bias variance decomposition becomes

$$\mathbb{E}_{\mathcal{D}}\left[E_{\text{ out}}\left(g^{(\mathcal{D})}\right)\right] = \sigma^2 + \text{bias} + \text{var}$$

(15 points)

2 VC Dimension

The VC dimension depends on the input space as well as $\mathcal{H}$. For a fixed $\mathcal{H}$, consider two input spaces $\mathcal{X}_1 \subseteq \mathcal{X}_2$. Show that the VC dimension of $\mathcal{H}$ with respect to input space $\mathcal{X}_1$ is at most the VC dimension of $\mathcal{H}$ with respect to input space $\mathcal{X}_2$. (15 points)

3 Comparing the VC-bounds

There are a number of bounds on the generalization error ϵ, all holding with probability at least $1 - \delta$.

(a) Original VC-bound:

$$\epsilon \leq \sqrt{\frac{8}{N} \ln \frac{4m_{\mathcal{H}}(2N)}{\delta}}\,.$$

(b) Rademancher Penalty Bound:

$$\epsilon \leq \sqrt{\frac{2\ln(2Nm_{\mathcal{H}}(N))}{N}} + \sqrt{\frac{2}{N}\ln\frac{1}{\delta}} + \frac{1}{N}.$$

(c) Parrondo and Van den Broek:

$$\epsilon \leq \sqrt{\frac{1}{N}\left(2\epsilon + \ln\frac{6m_{\mathcal{H}}(2N)}{\delta}\right)}.$$

(d) Devroye:

$$\epsilon \leq \sqrt{\frac{1}{2N}\left(4\epsilon(1+\epsilon) + \ln\frac{4m_{\mathcal{H}}(N^2)}{\delta}\right)}.$$

Note that (c) and (d) are implicit bounds in ϵ, therefore you are required to write out the explicit expressions. Fix $d_{VC} = 50$ and $\delta = 0.05$ and plot these bounds as a function of N. (Upper bound for $m_{\mathcal{H}}$ is needed as well.) Which is the best? Give your observations. (20 points)

4 Ridge regression

Ridge regression has two versions. One is the regularized version:

$$\min_{\theta} \quad \frac{1}{2}\|\mathbf{y} - \Phi\theta\|_2^2 + \frac{\mu}{2}\|\theta\|_2^2, \tag{1}$$

and the other is constrained version

$$\begin{aligned} \min_{\theta} \quad & \frac{1}{2}\|\mathbf{y} - \Phi\theta\|_2^2 \\ \text{s.t.} \quad & \|\theta\|_2^2 \leq C, \end{aligned} \tag{2}$$

with $C \geq 0$ and $\mu \geq 0$ are given parameters.

Hint: You can assume the optimal Lagrange multiplier λ is known. But you have to give the equation to determine λ.

Questions:

1. For any given $\mu \in [0, +\infty)$, find the optimal solution θ^{R1} to (1). (5 points)

2. Find the optimal solution θ^{R2} to (2) for any given $C \in [0, +\infty)$. (We've done the part for $C \geq \|\theta^{LS}\|_2^2$ in class using KKT conditions, where $\theta^{LS} = (\Phi^T\Phi)^{-1}\Phi^T\mathbf{y}$ be the LS solution.) (5 points)

3. For given $\mu \in [0, +\infty)$, you have the optimal solution θ^{R1} to (1). Now find the value of C, so that the optimal solution θ^{R2} to (2) is equivalent to θ^{R1}, i.e., $\theta^{R2} = \theta^{R1}$. (5 points)

4. For given $C \in [0, +\infty)$, you have the optimal solution θ^{R2} to (2). Now Find the μ value, so that (1) yields the same solution as (2), i.e., $\theta^{R1} = \theta^{R2}$.(5 points)

5 Nonlinear Transformation

A consumer price index (CPI) measures changes in the price level of a market basket of consumer goods and services purchased by households. The annual percentage change in a CPI is used as a measure of inflation. A CPI can be used to index (i.e., adjust for the effect of inflation) the real value of wages, salaries, pensions, for regulating prices and for deflating monetary magnitudes to show changes in real values. Generally, a CPI $\geq 3\%$ implies inflation, and a CPI $\geq 5\%$ indicates serious inflation. For a single item, the CPI is calculated by

$$\frac{CPI_2}{CPI_1} = \frac{Price_2}{Price_2}$$

where 1 means the comparison time period and CPI_1 is usually considered as an index of 100%. For multiple items, the overall CPI is given by

$$CPI = \frac{\sum_{i=1}^{n} CPI_i \times weight_i}{\sum_{i=1}^{n} weight_i}$$

where the $weight_i$s do not necessarily sum up to 1 or 100.

You are given the following data sets Table 1 and 2 about the monthly CPI of China[1].

Table 1: Monthly CPI of China, 2015. Increase as prior month

Time	CPI
January 2015 - December 2014	0.26%
February 2015 - January 2015	1.23%
March 2015 - February 2015	−0.52%
April 2015 - March 2015	−0.26%
May 2015 - April 2015	−0.17%
June 2015 - May 2015	0.00%
July 2015 - June 2015	0.35%
August 2015 - July 2015	0.52%
September 2015 - August 2015	0.09%
October 2015 - September 2015	−0.35%
November 2015 - October 2015	0.00%
December 2015 - November 2015	0.52%

Table 2: Monthly CPI of China, 2015. Increase as prior month

Time	CPI
January 2015 - January 2014	0.74%
February 2015 - February 2014	1.41%
March 2015 - March 2014	1.45%
April 2015 - April 2014	1.49%
May 2015 - May 2014	1.21%
June 2015 - June 2014	1.31%
July 2015 - July 2014	1.68%
August 2015 - August 2014	2.04%
September 2015 - September 2014	1.58%
October 2015 - October 2014	1.23%
November 2015 - November 2014	1.50%
December 2015 - December 2014	1.67%

Questions: (you can choose **either one** of the two tables above to do following homework):

[1] http://www.inflation.eu

1. Choose a regression model (e.g. linear regression, quadratic regression, cubic regression, or other nonlinear transformation you would like to use), write down your model. For example, if you choose linear model, it would be written as

$$CPI = \theta_0 + \theta_1 \times Time$$

where θ_0 and θ_1 are your parameters. (5 points)

2. Implement your regression in any programming language to determine your parameters, write down the values you find. (10 points)

3. Using LOOCV to calculate the cross-validation MSE and plot the 12 curves you obtain, (like I did in the class). (10 points)

4. Predict the CPI values for the next two time period, i.e., January 2016 and February 2016. (5 points)