# Coding Assignment: Consistency Bound of the Perceptron Classifiers

Due 2014/10/22 23:59:59

## 1 Goal

Given three training datasets $\{(x^{(t)}, r^{(t)})\}_{t=1}^{N}$ of different $N = 10$, 100, and 1000 respectively where $x^{(t)} \in \mathbb{R}$ is a scalar and $r^{(t)} \in \{1, -1\}$, your assignment is to 1) write the perceptron classifiers with three different lifting functions, $\Phi_1(x) = [x]$, $\Phi_2(x) = [x^2, x]$, and $\Phi_3(x) = [x^3, x^2, x]$, and 2) observe the gap between their generalization errors over the corresonding testing datasets $\{(x'^{(t)}, r'^{(t)})\}_{t=1}^{M}$ and the consistency bounds.

## 2 Steps

For each dataset and each lifting function:

1. Lift the input space to feature space.
2. Train a perceptron classifier $h := \arg\min_{g(\boldsymbol{x})=\boldsymbol{w}^\top \boldsymbol{x}+b} \Sigma_{t=1}^{N} 1(g(\Phi(\boldsymbol{x}^{(t)})) \neq r^{(t)})$ using the iterative training algorithm described below:

   (a) Initiate the $\boldsymbol{w}^{(0)}$ and $b^{(0)}$ randomly to obtain $h^{(0)}$;

   (b) At iteration $i = 0, 1, \cdots$, predict the label of an example $t$ using $h^{(i)}$:

   $$y^{(t)} := \operatorname{sgn}(h^{(i)}(\Phi(\boldsymbol{x}^{(t)}))) = \operatorname{sgn}(\boldsymbol{w}^{(i)}\Phi(\boldsymbol{x}^{(t)}) + b^{(i)})$$

   and obtain $h^{(i+1)}$ by:

   $$\boldsymbol{w}^{(i+1)} = \boldsymbol{w}^{(i)} + \eta \sum_{t=1}^{N} (r^{(t)} - y^{(t)})\Phi(\boldsymbol{x}^{(t)})$$

   and

   $$b^{(i+1)} = b^{(i)} + \eta \sum_{t=1}^{N} (r^{(t)} - y^{(t)})$$

   for some small learning rate $\eta > 0$;

(c) Repeate step (b) until either converagance ($\|\boldsymbol{w}^{(i+1)} - \boldsymbol{w}^{(i)}\| < \epsilon$) or the maximum iteration number (say 100) is reached.

3. Test your classifier using the provided corresponding testing dataset and get the generalization errors $R[h]$. You can't use any instances in the testing dataset when training your classifiers.

4. Given $\delta = 0.1$, plot the generalization error $R[h]$ and its consistency bound:

$$R[h^*] + 2\sqrt{\frac{32}{N}\left(VC(\mathcal{H})\log\frac{Ne}{VC(\mathcal{H})} + \log\frac{4}{\delta}\right)}.$$

## 3 Datasets

The $(x, r)$ pairs in the given datasets are generated using the following rules:

- $0.4N$ examples with label 1 in $[0, \frac{1}{9})$ uniformly
- $0.5N$ examples with label -1 in $[\frac{1}{9}, \frac{2}{9})$ uniformly
- $0.1N$ examples with label 1 in $[\frac{2}{9}, \frac{1}{3})$ uniformly

where $N$ is the dataset size. Use the above clues to figure out $R[h^*]$.

## 4 Submission Requirements

- A report (README.*) containing:
  - 3 figures, one for each lifting function. In each figure, there should be two lines telling how the generalization error and consistency bound various as $N$ becomes larger. The y-axis shows the values of errors and the x-axies shows the $N$ in the 10, 100, 1000 order.
  - How you implement the classifier
    - Discuss the figures mentioned above
    - Anything else worth mentioning
- Pack all the above with your solutions to Assignment 2.