Homework XX

MATH 591 Mathematics of Machine Learning
Fall 2019

due: 5pm Oct 24, submit on mycourses

Homework based on Lectures Oct 8-Oct 17

1. *PAC learning of hyper rectangles.* Mohri Chapter 2, question 2.2.

2. *PAC learning of intervals.* Mohri Chapter 2, question 2.9 (hint: do question 2.8 to warm up).

3. *Stochastic Gradient Descent (SGD) theory.* Suppose f is μ-strongly convex and L-smooth. Consider the abstract SGD iteration $x_{k+1} = x_k - h_k(\nabla f(x_k) + e_k)$, where e_k is random with mean zero and variance σ^2. Prove the following variant of the basic lemma,
$$\mathbb{E}[q(x_{k+1}) \mid x_k] \leq (1 - \mu h_k)q(x_k) + \frac{1}{2}h_k^2\sigma^2$$
provided $0 < h_k \leq 1/L$.

4. *Stochastic Gradient Descent (SGD) implementation.* Let $x_1, \ldots x_m$ be be random i.i.d. chosen uniformly on $[-1, 1]$. Consider the function $f(x) = \frac{1}{m}\sum_{i=1}^m (x - x_i)^2/2$. Take $m = 1000$

   (a) Find the best constants μ-convex and L-smooth for f. Find x^* and f^*.

   (b) Perform gradient descent $x_{k+1} = x_k - h_k(\nabla f(x_k))$ with $h_k = 1/2L$. Plot $\log(f(x_k) - f^*)$ against k.

   (c) Perform SGD $x_{k+1} = x_k - h_k \nabla_{mb} f(x_k)$ where ∇_{mb} is the gradient of a random mini-batch of f of size 100. (Scale the function by $1/100$ instead of $1/m$). Choose $h_k = 1/2L$ for $k = 1, \ldots 100$, then divide h by 5 and run for another 100 steps. Repeat this two more times. Plot the gap $f(x_k) - f^*$ as a function of k.

   (d) Repeat the last step, but now tune the time step in order to optimize f to the best possible. The results will be noisy, so you will need to plot the average values of f over 10 different runs. Compare the average over 10 runs of your tuned SGD to the original schedule.

5. *Convexity and Smoothness.* Let $f(x) = \|Mx - b\|^2/2$. Find the best μ-convexity and L-smoothness constants for $f(x)$ in terms of M and b.