Introduction to Machine Learning, Spring 2023
Homework 5

(Due Tues, May 23 at 11:59pm (CST))

April 29, 2023

1. [20 points] Consider the data $(X_1, Y_1), \ldots, (X_n, Y_n)$ where $X_i \in \mathbb{R}$ and $Y_i \in \mathbb{R}$. Inspired by the fact that $\mathbb{E}[Y \mid X = x] = \int yp(x, y)dy/p(x)$, define

$$\widehat{m}(x) = \frac{\int y\widehat{p}(x, y)dy}{\widehat{p}(x)}$$

where

$$\widehat{p}(x) = \frac{1}{n}\sum_i \frac{1}{h}K\left(\frac{X_i - x}{h}\right)$$

and

$$\widehat{p}(x, y) = \frac{1}{n}\sum_i \frac{1}{h^2}K\left(\frac{X_i - x}{h}\right)K\left(\frac{Y_i - y}{h}\right).$$

Assume that $\int K(u)du = 1$ and $\int uK(u)du = 0$. Show that $\widehat{m}(x)$ is exactly the kernel regression estimator $\frac{\sum K\left(\frac{x-X_i}{h}\right)Y_i}{\sum K\left(\frac{x-X_i}{h}\right)}$.

solution:

2. [30 points] Let f be differentiable, m-strongly convex, M-smooth and with minimizer x^* . In this exercise, we explore how to prove convergence in the function value difference $f\left(x^l\right) - f\left(x^*\right)$ for gradient descent with step size $\alpha = 1/M$.

   (a) Prove that:
   $$f\left(x^{l+1}\right) - f\left(x^*\right) \leq f\left(x^l\right) - f\left(x^*\right) - \frac{1}{2M}\left\|\nabla f\left(x^l\right)\right\|_2^2$$
   This shows that we have a descent method [10 points]
   solution:

   (b) Prove that:
   $$\frac{m}{M}\left(f\left(x^l\right) - f\left(x^*\right)\right) \leq \frac{1}{2M}\left\|\nabla f\left(x^l\right)\right\|_2^2$$
   [10 points]
   solution:

   (c) Conclude that:
   $$f\left(x^{l+1}\right) - f\left(x^*\right) \leq \left(1 - \frac{m}{M}\right)\left(f\left(x^l\right) - f\left(x^*\right)\right)$$
   This shows that we have geometric convergence with parameter $1 - \frac{m}{M}$[10 points]
   solution: