ELE 535: Machine Learning and Pattern Recognition
Homework 4

Zachary Hervieux-Moore

Monday 15$^{\text{th}}$ October, 2018

Exercise 1: Determine general sufficient conditions (if any exist) under which the indicated function f is convex.

a) $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = |x|$.

b) $f : (0, \infty) \to \mathbb{R}$ with $f(x) = x \ln(x)$.

c) $f : \mathbb{R}^n \to \mathbb{R}$ with $f(x) = (x^T Q x)^3$. Here $Q \in \mathbb{R}^{n\times n}$ is symmetric PSD.

d) $f : \mathbb{R}^n \to \mathbb{R}$ with $f(x) = 1 + e^{\sum_{i=1}^n |x_i|^3}$.

e) For $x \in \mathcal{C} = \{x \in \mathbb{R}^n : x_i > 0, i = 1, \dots, n\}$ let $\ln(x) = [\ln(x_i)] \in \mathbb{R}^n$ and define $f(x) = x^T \ln(x)$

Answer:

a) Simply applying the triangle inequality we get

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= |\lambda x + (1-\lambda)y| \\ \leq \lambda|x| + (1-\lambda)|y| &= \lambda f(x) + (1-\lambda)f(y) \end{aligned}$$

Thus, it is convex.

b) Here, we take the second derivative since it is continuous on the domain,

$$f''(x) = \frac{1}{x} > 0 \quad \forall x \in (0, \infty)$$

As this is always positive, then we have that f is strongly convex.

c) Here, we use theorem 7.3.2e) that states that $h(x) = g(f(x))$ is convex if g is convex and non decreasing on the range of f. Here we have $f(x) = x^T Q x$ whose range is $[0, \infty)$ as Q is PSD and $g(x) = x^3$. Thus, on $[0, \infty)$, we have that $g'(x) = 2x^2 \geq 0$ and $g''(x) = 6x \geq 0$ which means that g is on decreasing and convex respectively. Thus, we have shown that the original function is convex.

d) Again, we use the same theorem as in the previous part. First $g(x) = 1 + e^x$ and $f(x) = (\sum_{i=1}^n |x_i|)^3$. The range of $f(x)$ is $[0, \infty)$. We also have $g'(x) = e^x$ and $g''(x) = e^x$ which are both always positive and so non decreasing and convex. This shows that the original function is convex.

e) We have that $f(x) = x^T \ln(x)$ which is just a short form for

$$[f(x)]_i = x_i \ln(x_i)$$

Thus, we have

$$[\nabla f(x)]_i = 1 + \ln(x_i)$$
$$[Hf(x)]_{ii} = \frac{1}{x_i} \text{ and } [Hf(x)]_{ij} = 0 \text{ for } i \neq j$$

Exercise 2: You want to learn an unknown function $f : [0, 1] \to \mathbb{R}$ using a set of noisy measurements (x_j, y_j), with $y_j = f(x_j) + \epsilon_j$, $j = 1, \dots, m$. Your plan is to approximate $f(\cdot)$ by a Fourier series on $[0, 1]$ with $q \in \mathbb{N}$ terms:

$$f_q(x) = \frac{a_0}{2} + \sum_{k=1}^{q} a_k \cos(2\pi kx) + b_k \sin(2\pi kx).$$

To control the smoothness of $f_q(\cdot)$, you also decide to penalize the size of the coefficients a_k, b_k more heavily as k increases.

a) Formulate the above problem as a regularized regression problem.

b) For $q = 2$, display the regression matrix, the label vector y, and the regularization term.

c) Comment briefly on how to select q.

Answer:

Exercise 3: Let $D \in \mathbb{R}^{n\times n}$ be diagonal with nonnegative diagonal entries and consider the problem:

$$\min_{x\in\mathbb{R}^n} \|x-y\|_2^2 + \lambda\|Dx\|_2^2$$

This problem seeks to best approximate $y \in \mathbb{R}^n$ with a nonuniform penalty for large entries in x.

a) Solve this problem using the solution of ridge regression.

b) Show that the objective function is separable into a sum of decoupled terms. Show that this decomposes the problem into n independent scalar problems.

c) Finc the solution of each scalar problem.

d) By putting these scalar solutions together, find and interpret the solution to the original problem.

Answer:

Exercise 4: Let $X \in \mathbb{R}^{n \times m}$ and $y \in \mathbb{R}^m$ be given, and $\lambda > 0$. Consider the problem

$$w^* = \arg\min_{w \in \mathbb{R}^n} \|y - X^T w\|_2^2 + \lambda \|w\|_2^2$$

From the notes we know that there exists a unique solution w^* and that $w^* \in \mathcal{R}(X)$. Using the above and these two results, show that $w^* = X(X^T X + \lambda I_m)^{-1} y$.

Answer:

Exercise 5: One form of regularized least squares can be posed as:

$$w^* = \arg\min_{w \in \mathbb{R}^n} \|Fw - y\|_2^2 + \lambda \|Gw - g\|_2^2$$

where $F \in \mathbb{R}^{m \times n}$, $y \in \mathbb{R}^m$, $G \in \mathbb{R}^{k \times n}$, $g \in \mathbb{R}^k$, and $\lambda > 0$.

a) Show that a sufficient condition for the above to have a unique solution is that $\mathrm{rank}(G) = n$.

b) Show that a necessary and sufficient condition is that $\mathcal{N}(F) \cap \mathcal{N}(G) = \mathbf{0}$.

Answer: