Homework #2B

CSE 546: Machine Learning
Cassia Cai
October 31, 2022

B1. *[6 points]* For any $x \in \mathbb{R}^n$, define the following norms: $\|x\|_1 = \sum_{i=1}^n |x_i|$, $\|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$, $\|x\|_\infty := \lim_{p\to\infty} \|x\|_p = \max_{i=1,\ldots,n} |x_i|$. Show that $\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1$.

$$||x||_1^2 = (\sum_{i=1}^n |x_i| + 2\sum_{i\neq j} |x_i||x_j|) \geq \sum_{i=1}^n |x_i|^2 = ||x||_2^2 \tag{1}$$

Thus, we see that:

$$||x||_1 \geq ||x||_2 \tag{2}$$

$$||x||_{\text{inf}}^2 = (max_{i=1,\ldots n}|x_i|)^2 = max_{i=1,\ldots n}|x_i|^2 \leq \sum_{i=1}^n |x_i|^2 = ||x||_2^2 \tag{3}$$

Thus, we see that:

$$||x||_2 \geq ||x||_{\text{inf}} \tag{4}$$

Together, we have:

$$||x||_1 \geq ||x||_2 \geq ||x||_{\text{inf}} \tag{5}$$

This is the same as: $\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1$.

B2. For $i = 1, \ldots, n$ let $\ell_i(w)$ be convex functions over $w \in \mathbb{R}^d$ (e.g., $\ell_i(w) = (y_i - w^\top x_i)^2$), $\|\cdot\|$ is any norm, and $\lambda > 0$.

a. *[3 points]* Show that

$$\sum_{i=1}^{n} \ell_i(w) + \lambda \|w\|$$

is convex over $w \in \mathbb{R}^d$ (Hint: Show that if f, g are convex functions, then $f(x) + g(x)$ is also convex.)

We can first show that f(x) is convex. $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$ where λ is in the range of 0 and 1.

$$f(\lambda x + (1-\lambda)y) + g(\lambda x + (1-\lambda)y) \leq f(\lambda x) + f((1-\lambda)y) + g(\lambda x) + g((1-\lambda)y) \tag{6}$$

$$f(\lambda x) + f((1-\lambda)y) + g(\lambda x) + g((1-\lambda)y) = \lambda(f(x) + g(x)) + (1-\lambda)((f(y) + g(y)) \tag{7}$$

Then, we have shown that $f(x) + g(x)$ is also convex. Here, we are considering a sum of n convex functions. Since we know that $f_i(x)$ is convex (from above), and that $f_1(x) + f_2(x)$ is also convex, then we know that $f_1(x) + f_2(x) + f_3(x)$ is also convex. Therefore, $\sum_{i=1}^{n} \ell_i(w) + \lambda\|w\|$ is also convex since it is a sum of convex functions (as $\|w\|$ is also convex).

b. *[1 point]* Explain in one sentence why we prefer to use loss functions and regularized loss functions that are convex.

Convex loss functions and convex regularized loss functions are preferred because the property of convex-ness guarantees that our local minimum is the global minimum.

B3. Here, we continue the previous A problem and finish the proof. DID NOT ATTEMPT!

a. *[3 points]* By induction on k, show that $||w_k - w^*|| \leq \rho^k ||w_0 - w^*||$ for any positive integer k.

b. *[5 points]* Show that if $\alpha = \frac{1}{h+H}$, then $0 \leq \rho < 1$. Then show that for any error term $\varepsilon > 0$, if

$$k > \log_{1/\rho}\left(\frac{||w_0 - w^*||}{\varepsilon}\right)$$

then $||w_k - w^*|| < \varepsilon$. This shows that w_k gets arbitrarily close to w^* as the number of iterations increases, meaning that the gradient descent method converges to the optimal solution.