Homework #2B

Spring 2020, CSE 446/546: Machine Learning
Prof. Kevin Jamieson and Prof. Jamie Morgenstern
Due: 5/12/20 11:59 PM

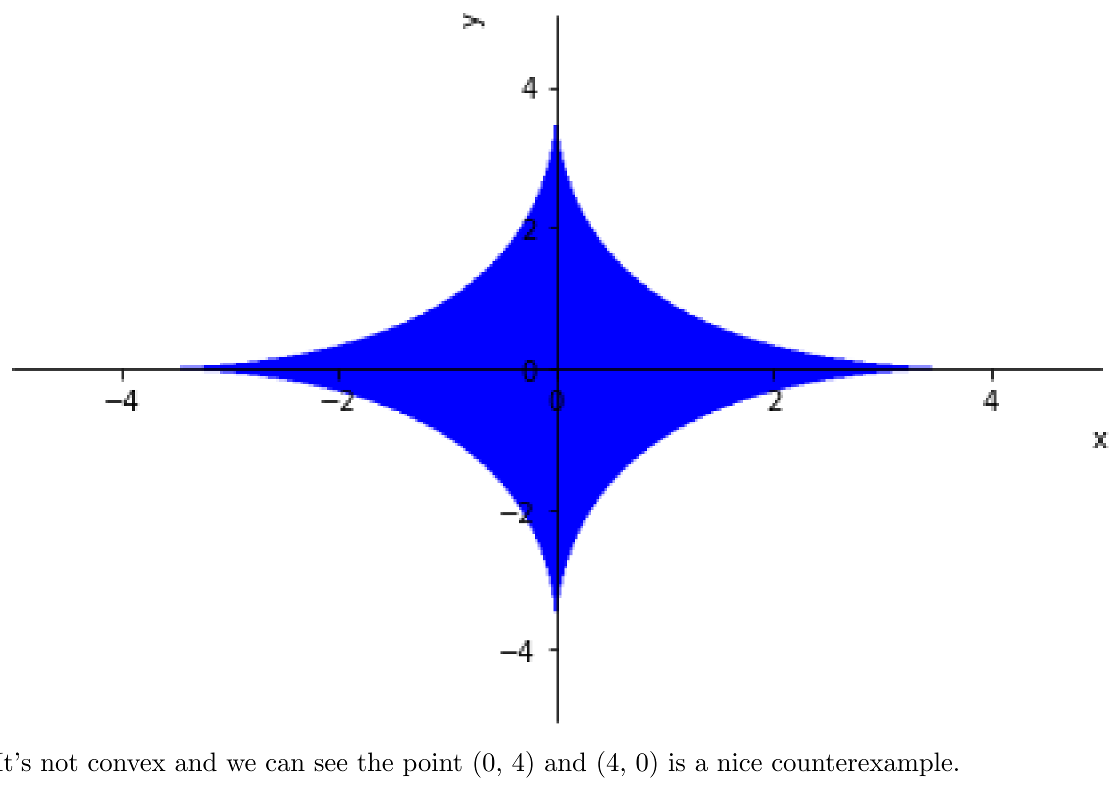
1 Convexity and Norms

B.1
Firstly, for the first inequality:
$||x||_\infty = max_{i=1,...,n}|x_i| \leq \sqrt{(max_{i=1,...,n}|x_i|)^2} \leq \sqrt{\sum_1^n (|x_i|)^2} = ||x||_2$
and then for the second inequality:
$(||x||_2)^2 \leq (||x||_1)^2$ since when we check
$(||x||_2)^2 = \sum_1^n (|x_i|)^2$ and $(||x||_1)^2 = (\sum_1^n |x_i|)^2$
we find that $(||x||_1)^2$ include some more non-negative interaction terms. so this inequality holds.

B.2

a. For any $\lambda \in [0, 1]$ and $x, y \in R^n$
   $f(\lambda x + (1-\lambda)y) = ||\lambda x + (1-\lambda)y|| \leq ||\lambda x|| + ||(1-\lambda)y|| = \lambda||x|| + (1-\lambda)||y|| = \lambda f(x) + (1-\lambda)f(y)$
   So f(x) is a convex function.

b. For x_1, x_2 in this set
   $\lambda||x_1|| + (1-\lambda)||x_2|| \leq$1 since$||x_1|| \leq 1$ and $||x_2|| \leq 1$.



c. It's not convex and we can see the point (0, 4) and (4, 0) is a nice counterexample.

B.3

a. Set $h(x) = f(x) + g(x)$ where f(x) and g(x) are both convex functions.
$h(\lambda x + (1-\lambda)y) = f(\lambda x + (1-\lambda)y) + g(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) + \lambda g(x) + (1-\lambda)g(y)$
$= \lambda(f(x) + g(x)) + (1-\lambda)(f(x) + g(x))$
$= \lambda h(x) + (1-\lambda)h(x)$
so h(x) is also a convex function and here we can see each $\ell_i(w) and ||w||$ are convex function and due to mathematical induction:
$\sum_1^n \ell_i(w) + \lambda||w||$ is a convex function

b. Because a local minimum of a convex function is a global minimum so it's easier to find global optimum.

B.4

a. $\frac{dlog(\frac{exp(\mathbf{w}^{(\ell)}\cdot\mathbf{x}_i)}{\sum_{j=1}^k exp(\mathbf{w}^{(j)}\cdot\mathbf{x}_i)})}{dW}$
$= \frac{\sum_{j=1}^k exp(\mathbf{w}^{(j)}\cdot\mathbf{x}_i)}{exp(\mathbf{w}^{(\ell)}\cdot\mathbf{x}_i)} \cdot \frac{\mathbf{x}_i exp(\mathbf{w}^{(\ell)}\cdot\mathbf{x}_i)\sum_{j=1}^k exp(\mathbf{w}^{(j)}\cdot\mathbf{x}_i) - exp(\mathbf{w}^{(\ell)}\cdot\mathbf{x}_i)\sum_{j=1}^k \mathbf{x}_i exp(\mathbf{w}^{(j)}\cdot\mathbf{x}_i)}{(\sum_{j=1}^k exp(\mathbf{w}^{(j)}\cdot\mathbf{x}_i))^2}$
$= \frac{\mathbf{x}_i \sum_{j=1}^k exp(\mathbf{w}^{(j)}\cdot\mathbf{x}_i) - \sum_{j=1}^k \mathbf{x}_i exp(\mathbf{w}^{(j)}\cdot\mathbf{x}_i)}{\sum_{j=1}^k exp(\mathbf{w}^{(j)}\cdot\mathbf{x}_i)}$
So
$\nabla_w L(W) = -\sum_{i=1}^n \sum_{\ell=1}^k 1\{y_i = \ell\} \frac{\mathbf{x}_i \sum_{j=1}^k exp(\mathbf{w}^{(j)}\cdot\mathbf{x}_i) - \sum_{j=1}^k \mathbf{x}_i exp(\mathbf{w}^{(j)}\cdot\mathbf{x}_i)}{\sum_{j=1}^k exp(\mathbf{w}^{(j)}\cdot\mathbf{x}_i)}$
$= -\sum_{i=1}^n x_i(y_i)^T - x_i(softmax(W^T x_i))^T$
so the equality holds

b. $J(W) = \frac{1}{2}\sum_{i=1}^n (y_i - W^T x_i)^2$
$\nabla_w J(w,b) = -\sum_{i=1}^n x_i(y_i - W^T x_i)$
$= -\sum_{i=1}^n x_i(y_i - (\tilde{y_i})^{(W)})^T$