MATH 462 ASSIGNMENT 2
DRAFT VERSION October 7, 2022

2.1. **Instructions.** Refer to notes and references on `https://adam-oberman.github.io/Math462/`. Submit your solutions on MyCourses course page. Math exercises should be *handwritten*. You can get help from other students, but you should do the write up yourself. Coding exercises: export a PDF of the plots required.

2.2. **Calculus of Softmax.** Refer to course notes.

Exercise 2.1. Sketches of softmax functions.

(a) Make a sketch of $\sigma(x) = \frac{1}{1+\exp(-x)}$, also sketch $\sigma'(x)$

(b) Sketch a contour plot of $\mathrm{LSE}(x) = \log(\exp(x_1) + \exp(x_2))$. Indicate the gradient on the plot using arrows.

Exercise 2.2. Show that

(a) the odds ratio function $r(p) = p/(1-p)$ is invertible on $\mathbb{R}^+$ with inverse $p(r) = r/(r+1)$.

(b) Show that σ and logit are inverses. (Hint: you can write $\sigma(x) = p(\exp(x))$, and $\mathrm{logit}(p) = \log(r(p))$).

Exercise 2.3. Verify the following properties for $\sigma(x)$.

(a) $1 - \sigma(x) = \sigma(-x)$

(b) $\sigma'(x) = \sigma(x)(1 - \sigma(x))$

Exercise 2.4. Let $x \in \mathbb{R}^K$.

(a) Prove that $\max(x) \le \mathrm{LSE}(x) \le \max(x) + \log(K)$,

(b) Consider the scaled version of the function, $f(x,t) = \frac{1}{t}\mathrm{LSE}(tx)$. Prove that $f(x,t) \to \max(x)$ as $t \to \infty$. (Hint, use the previous part). Interpret this result in terms of the function name soft-arg-max.

Exercise 2.5. Updated When $x \in \mathbb{R}$, we defined $\sigma(x) = \frac{1}{1+\exp(-x)}$. When $x \in \mathbb{R}^2$, we defined $\sigma(x) = (e^{x_1}, e^{x_2})/(e^{x_1} + e^{x_2})$. Given $x \in \mathbb{R}$, find $(x_1, x_2) \in \mathbb{R}^2$ so that

$$\frac{e^{x_1}}{(e^{x_1} + e^{x_2})} = \frac{1}{1 + \exp(-x)}$$

Remark: the point of this question is to show that the one dimensional function $\sigma(x)$ is a special case of the vector version of $\sigma(x)$.

2.3. **Coding for k-means.**

Exercise 2.6. Use the code presented in class (see course page) for the k-means algorithm.

(1) Generate 3 blobs, and find a 3-means cluster. On the same 3 blobs, do a 6-means cluster. Give examples of a nearly optimal cluster, along with a suboptimal cluster and compare the final loss of each.

(2) Generate the uniform distribution on the 4 by 2 rectangle with vertices $(\pm 2, \pm 1)$. Use the appropriate numpy uniform random number generator. Show the clustering you obtain. ~~Show two different ways to cluster with $k = 2$. (You can force the clustering by modifying the code to take as input the initial means. Or, you can just run the code till you get lucky). On your plot, of the losses as a function of t, compare the losses the two cases.~~ (Second part removed, too hard to get the other cluster).

2.4. **Analytic Geometry and Covariance matrices.**

Exercise 2.7. Exercise 3.5 from MML book. Write down a system of linear equations involving the problem inputs. You can use software to solve the system.

Exercise 2.8. Exercise 3.6 from MML book. Write down a system of linear equations involving the problem inputs. You can use software to solve the system.

Exercise 2.9 (Inner products). (a) Given full rank (invertible) $n \times n$ matrix M, show that $\langle x, y\rangle_M = (Mx)^\top (My)$ defines an inner product on $\mathbb{R}^n$. What goes wrong if the matrix has a non-trivial null-space?

(b) Given an example of a norm on $\mathbb{R}^n$ which does not come from an inner product.

Exercise 2.10 (Covariance Matrix). Let $n = 2$. Find the covariance matrix of S^m.

(a) $S^m = \{(1,1), (-1,-1), (1,0), (-1,0), (-1,1), (1,-1), (0,1), (0,-1)\}$

(b) $S^m = \{(t,t), (-t,-t), (1,0), (-1,0), (-1,1), (1,-1), (0,1), (0,-1)\}$, for any $t \in R$.

Hint: recall for a vector $v = (a,b)$, $v^\top v = a^2 + b^2$ is a scalar, and $vv^\top = \begin{bmatrix} a^2 & ab \\ ab & b^2 \end{bmatrix}$ is a matrix.

Exercise 2.11 (Covariance Matrix Theory).

(a) Prove that the covariance matrix, C, is symmetric and non-negative definite, meaning $x^\top Cx \geq 0$ for all x. Assuming that the matrix is invertible, prove it is (strictly) positive definite.

(b) Given $S^m = \{x_1, \ldots x_m\}$ with $x_i \in \mathbb{R}^n$. Let C be the covariance matrix of S^m. Let M be an invertible $n \times n$ matrix Define $\tilde{S}^m = \{Mx_1, \ldots Mx_n\}$. Find a formula for $\tilde{C}$, the covariance matrix of $\tilde{S}^m$, in terms of C and M.