AMATH 515 Homework #1 - Vinsensius
Due: Monday, January 23th, by 11:59 pm

Number 1

Let $g : \mathbb{R}^m \to \mathbb{R}$ is a twice differentiable function, $A \in \mathbb{R}^{m\times n}$ any matrix, and h is the composition $g(Ax)$, then we have two simple generalizations of the chain rule that combine linear algebra with calculus:

$$\nabla h(x) = A^T \nabla g(Ax)$$

and

$$\nabla^2 h(x) = A^T \nabla^2 g(Ax) A.$$

(a) Show what happens when you apply the above chain rules to the special case

$$h(x) = g(a^T x)$$

where a is a vector.

(b) Compute the gradient and hessian of the regularized logistic regression objective:

$$\left(\sum_{i=1}^{n} \log(1 + \exp(a_i^T x)) - b^T Ax\right) + \lambda\|x\|^2$$

where a_i denote the rows of A.

(c) Compute the gradient and hessian of the regularized poisson regression objective:

$$\left(\sum_{i=1}^{n} \exp(a_i^T x) - b^T Ax\right) + \lambda\|x\|^2$$

where a_i denote the rows of A.

(d) Compute the gradient and hessian of the regularized 'concordant' regression objective

$$\|Ax - b\|_2 + \lambda\|x\|_2.$$

Give conditions on a point x that ensure the gradient and Hessian exist at x.

(a)

$$\nabla h(x) = a \nabla g(a^T x)$$

$$\nabla^2 h(x) = a\nabla^2 g(a^T x)a^T$$

(b) Let $h(x) = \left(\sum_{i=1}^n \log(1+\exp(a_i^T x)) - b^T Ax\right) + \lambda\|x\|^2$

We can look at $f(y) = \log(1+\exp(y))$, we find that $f'(y) = \frac{\exp(y)}{(1+\exp(y))}$ and $f''(y) = \frac{\exp(y)}{(1+\exp(y))^2}$. Using this we can use composition rule with $a^T x$ like in previous part to calculate the gradient and the hessian.

$$\begin{aligned}
\nabla h(x) &= \sum_{i=1}^n \nabla[\log(1+\exp(a_i^T x))] - \nabla(b^T Ax) + \lambda\nabla(\|x\|^2) \\
&= \sum_{i=1}^n \frac{1}{(1+\exp(a_i^T x)}\nabla(1+\exp(a_i^T x)) - A^T b + \lambda\nabla(\|x\|^2) \\
&= \sum_{i=1}^n \frac{1}{(1+\exp(a_i^T x)} a_i \exp(a_i^T x) - A^T b + \lambda\nabla(\|x\|^2)
\end{aligned}$$

$$\begin{aligned}
\nabla^2 h(x) &= \sum_{i=1}^n \nabla^2[\log(1+\exp(a_i^T x))] - \nabla^2(b^T Ax) + \lambda\nabla^2(\|x\|^2) \\
&= \sum_{i=1}^n \nabla\left[\frac{1}{(1+\exp(a_i^T x)} a_i \exp(a_i^T x)\right] + 2\lambda\mathbf{I} \\
&= \sum_{i=1}^n \frac{1}{(1+\exp(a_i^T x)^2} a_i \exp(a_i^T x) a_i^T + 2\lambda\mathbf{I}
\end{aligned}$$

(c) Let $h(x) = \left(\sum_{i=1}^n \exp(a_i^T x) - b^T Ax\right) + \lambda\|x\|^2$

$$\begin{aligned}
\nabla h(x) &= \sum_{i=1}^n \nabla[\exp(a_i^T x)] - \nabla(b^T Ax) + \lambda\nabla(\|x\|^2) \\
&= \sum_{i=1}^n a_i \exp(a_i^T x) - A^T b + 2\lambda x
\end{aligned}$$

$$\begin{aligned}
\nabla^2 h(x) &= \sum_{i=1}^n \nabla^2[\exp(a_i^T x)] - \nabla^2(b^T Ax) + \lambda\nabla^2(\|x\|^2) \\
&= \sum_{i=1}^n a_i \exp(a_i^T x) a_i^T + 2\lambda\mathbf{I}
\end{aligned}$$

(d)

$$\nabla\left(||Ax-b||_2+\lambda||x||_2\right) = \nabla\left[\left(\sum_{i=0}^{m}(a_i^Tx-b_i)^2\right)^{0.5}\right] - \lambda\nabla\left[\left(\sum_{i=0}^{n}x_i^2\right)^{0.5}\right]$$

We will look at the $\nabla\|x\|$

$$\begin{aligned}
\nabla\|x\| &= \nabla\left[\left(\sum_{i=0}^{n}x_i^2\right)^{0.5}\right] \\
&= \frac{1}{2}\frac{1}{\left(\sum_{i=0}^{n}x_i^2\right)^{-0.5}}\nabla\left(\sum_{i=0}^{n}x_i^2\right) \\
&= \frac{x}{\|x\|_2}
\end{aligned}$$

Then, we will look at the $\nabla\|Ax-b\|$

$$\begin{aligned}
\nabla\|Ax-b\| &= \nabla\left[\left(\sum_{i=0}^{m}(a_i^Tx-b_i)^2\right)^{0.5}\right] \\
&= \frac{1}{2}\frac{1}{\left(\sum_{i=0}^{m}(a_i^Tx-b_i)^2\right)^{-0.5}}\nabla\left(\sum_{i=0}^{n}(a_i^Tx-b_i)^2\right) \\
&= \frac{1}{\left(\sum_{i=0}^{m}(a_i^Tx-b_i)^2\right)^{-0.5}}\sum_{i=0}^{n}a_i(a_i^Tx-b_i) \\
&= \frac{A^T(Ax-b)}{\|Ax-b\|_2}
\end{aligned}$$

Combining them we get

$$\nabla\left(||Ax-b||_2+\lambda||x||_2\right) = \frac{A^T(Ax-b)}{\|Ax-b\|_2} + \lambda\frac{x}{\|x\|_2}$$

Now we are looking at the hessian,

$$\nabla^2\left(||Ax-b||_2+\lambda||x||_2\right) = \nabla^2\left[\left(\sum_{i=0}^{m}(a_i^Tx-b_i)^2\right)^{0.5}\right] - \lambda\nabla^2\left[\left(\sum_{i=0}^{n}x_i^2\right)^{0.5}\right]$$

We will be looking at $\nabla^2\|x\|_2$

$$
\begin{aligned}
\nabla^2||x||_2 &= \nabla^2\left[\left(\sum_{i=0}^{n} x_i^2\right)^{0.5}\right] \\
&= \nabla\left[\frac{x}{\|x\|_2}\right] \\
&= \frac{\nabla(x)}{\|x\|_2} + x\left(\nabla\left(\sum_{i=0}^{n} x_i^2\right)^{-0.5}\right) \\
&= \frac{\mathbf{I}}{\|x\|_2} - \frac{1}{2}x\left(\sum_{i=0}^{n} x_i^2\right)^{-1.5}\left[\nabla\left[\sum_{i=0}^{n} x_i^2\right]\right]^T \\
&= \frac{\mathbf{I}}{\|x\|_2} + \frac{xx^T}{\|x\|_2^3}
\end{aligned}
$$

Now, we will be looking at $\nabla^2\|Ax-b\|_2$

$$
\begin{aligned}
\nabla^2\|Ax-b\|_2 &= \nabla^2\left[\left(\sum_{i=0}^{m}(a_i^Tx-b_i)^2\right)^{0.5}\right] \\
&= \nabla\left[\frac{A^T(Ax-b)}{\|Ax-b\|_2}\right] \\
&= \frac{\nabla\left[A^T(Ax-b)\right]}{\|Ax-b\|_2} + \left(A^T(Ax-b)\right)\left(\nabla\left(\sum_{i=0}^{m}(a_i^Tx-b_i)^2\right)^{-0.5}\right) \\
&= \frac{A^TA}{\|Ax-b\|_2} - \frac{1}{2}\left(A^T(Ax-b)\right)\left(\sum_{i=0}^{n}(a_i^Tx-b_i)^2\right)^{-1.5}\left[\nabla\left[\sum_{i=0}^{n}(a_i^Tx-b_i)^2\right]\right]^T \\
&= \frac{A^TA}{\|x\|_2} + \frac{A^T(Ax-b)(Ax-b)^TA}{\|x\|_2^3}
\end{aligned}
$$

Combining them we get

$$
\nabla^2\left(||Ax-b||_2 + \lambda||x||_2\right) = \frac{A^TA}{||Ax-b||_2} + \frac{A^T(Ax-b)(Ax-b)^TA}{||Ax-b||_2^3} + \lambda\left[\frac{\mathbf{I}}{\|x\|_2} + \frac{xx^T}{\|x\|_2^3}\right]
$$

The condition for the $\|Ax-b\|_2 \neq 0$ and $\|x\|_2 \neq 0$

Number 2

Show that each of the following functions is convex.

(a) Indicator function to a convex set: $\delta_C(x) = \begin{cases} 0 & \text{if} \quad x \in C \\ \infty & \text{if} \quad x \notin C. \end{cases}$

(b) Support function to any set:

$$\sigma_C(x) = \sup_{c \in C} c^T x.$$

(c) Any norm (see Chapter 1 for definition of a norm)

(a) Using the algebraic expression $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$, there are 3 cases:

  (i) Suppose $x, y \in C$. Since x,y are in convex set, the convex combination of x,y, i.e. $\lambda x + (1-\lambda)y \in C$. So, the LHS of the inequality is 0 while the RHS of the inequality is 0. So, the convexity holds for this case.

  (ii) Without loss of generality, suppose $x \in C$, $y \notin C$. On the RHS, we will have ∞. So, the inequality becomes $f(\lambda x + (1-\lambda)y) \leq \infty$. This inequality holds since any function value from LHS will be $\leq \infty$.

  (iii) Suppose $x, y \notin C$. Since both are not in the convex set, LHS and RHS results in ∞. The inequality will still holds because $\infty \leq \infty$.

  Thus, the indicator function to convex set is convex.

(b) We want to show $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ through the function. So, we start with $\sigma_A(\lambda x + (1-\lambda)y)$.

$$\begin{aligned}
\sigma_A(\lambda x + (1-\lambda)y) &= \sup_{c \in A} c^T(\lambda x + (1-\lambda)y) \\
&\leq \sup_{c \in A} c^T(\lambda x) + \sup_{c \in A} c^T(1-\lambda)y) \quad \text{(Property of sup)} \\
&= \lambda \sup_{c \in A} c^T x + (1-\lambda) \sup_{c \in A} c^T y
\end{aligned}$$

  By comparison, the inequality is in the same form for algebraic definition of convex function, thus, the support function to any set is convex.

(c) We want to show $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ through the norm. Let norm be function f.

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &\leq f(\lambda x) + f((1-\lambda)y)(\text{ triangle inequality property of norm}) \\ &= \lambda f(x) + (1-\lambda)f(y)(\text{ scalar multiplication property of norm}) \end{aligned}$$

Thus, any norm is convex.

Number 3

Convexity and composition rules. Suppose that f and g are $\mathbb{C}^2$ functions from $\mathbb{R}$ to $\mathbb{R}$, with $h = f \circ g$ their composition, defined by $h(x) = f(g(x))$.

(a) If f and g are convex, show it is possible for h to be nonconvex (give an example). Give additional conditions that ensure the composition is convex.

(b) If f is convex and g is concave, what additional hypothesis that guarantees h is convex?

(c) Show that if $f : \mathbb{R}^m \to \mathbb{R}$ is convex and $g : \mathbb{R}^n \to \mathbb{R}^m$ affine, then h is convex.

(d) Show that the following functions are convex:

(i) Logistic regression objective: $\sum_{i=1}^n \log(1 + \exp(a_i^T x)) - b^T Ax$

(ii) Poisson regression objective: $\sum_{i=1}^n \exp(a_i^T x) - b^T Ax$.

(a) Suppose $f(x) = x^2$, $g(x) = x^2 - 1$. The resulting $h(x) = x^4 - 2x^2 + 1$ is not convex since $h''(x) \ngeq 0$. Additional hypothesis is $h''(x) \geq 0$.

(b) So, $h''(x) = f'(g(x))g''(x) + f''(g(x))(g'(x)^2)$. To make $h''(x) \geq 0$, f needs to be nonincreasing.

(c) The scalar equivalent is $h''(x) = f'(g(x))g''(x) + f''(g(x))(g'(x)^2)$. We know that $\nabla^2 g(x) = 0$ since it is affine function. So, $\nabla^2 h(x) = A^T \nabla^2 f(Ax + b)A$, from number 1 definition and $\nabla^2 h(x) \geq 0$ since f is convex and $A^T A \geq 0$ since it is equivalent of saying $y(x)^2$ on scalar function. So, $\nabla^2 h(x)$ is convex.

(d) (i) To prove this, we will look at the summation function only because when we do hessian to check convexity, the term $b^T Ax$ goes to 0. We will start by looking function under the summation $h(x) = \log(1 + \exp(a^T x))$ and prove that it is convex so that we can use the fact that the summation of convex functions is convex. We can think $h(x)$ as composition of a function, $f(x) = \log(1+\exp(x))$ with an affine function, $g(x) = a^T x$. To prove $h(x)$ is convex we can use result from 3(c). So, we need to prove that $f(x)$ to be convex to prove $h(x)$ is convex. We can look at $f''(x)$

$$f''(x) = \frac{\exp(x)}{(1 + \exp(x))^2}$$

By looking at the expression of the expression we know that $f''(x) > 0$ since $\exp(x) > 0$ and $\frac{1}{(1+\exp(x))^2} > 0$. Thus, $f(x)$ is convex. So, by looking at the result of 3(c), we can infer that $h(x)$ is convex since this is the special case of 3(c) since $g : \mathbb{R}^n \to \mathbb{R}$ and $f : \mathbb{R}^m \to \mathbb{R}$ so $h(x)$ is convex. So, the summation of convex functions $h_i(x)$ will be convex since the summation inequality $h_i(\lambda x+(1-\lambda)y) \leq \lambda h_i(x) + (1-\lambda)h_i(y)$ still holds. Thus, logistic regression objective is convex.

(ii) We can take a look of hessian of poisson regreesion objective to check its convexity.

$$
\begin{aligned}
\nabla^2 \left[\sum_{i=1}^{n} \exp(a_i^T x) - b^T A x \right] &= \sum_{i=1}^{n} a_i \exp(a_i^T x) a_i^T \quad \text{(From 1(c))} \\
&= A^T A \sum_{i=1}^{n} \exp(a_i^T x) \quad \text{(Vectorized form)}
\end{aligned}
$$

By looking at the expression, we know that $A^T A$ is positive semidefinite because the eigenvalues of $A^T A$ are the square of singular values of A corresponding to right singular eigenvector of A. Since $\exp(a_i^T x) > 0$, the $\sum_{i=1}^{n} \exp(a_i^T x) > 0$. Thus, the hessian of the poisson regression objective is ≥ 0 and it is convex.

Number 4

A function f is *strictly convex* if

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y), \quad \lambda \in (0,1).$$

(a) Give an example of a strictly convex function that does not have a minimizer.

(b) Show that a sum of a strictly convex function and a convex function is strictly convex.

(c) Characterize all solutions to the problem

$$\min_x \frac{1}{2}\|Ax - b\|^2$$

(a) $f(x) = \exp(x)$ is strictly convex because $f''(x) > 0$ but it does not minimizer because it does not have stationary point.

(b) We can do the summation of both inequality

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$$
$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

We get

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$$

The resulting inequality sign $<$ because when addition of $\leq$ and $<$ will be the intersection of the two inequality which is $<$. Thus, the sum of strictly convex and convex function is strictly convex.

(c) To find the solution of the problem, we will find the gradient of the function $\frac{1}{2}\|Ax-b\|^2$ and set it to 0,

$$\nabla \left[\frac{1}{2}\|Ax - b\|^2\right] = 0$$
$$A^T(Ax - b) = 0$$
$$A^T Ax = A^T b$$

Then, we can find x by inverting $(A^T A)$ to the other side so we get $x = (A^T A)^{-1}A^T b$. If A has full rank, x will be unique since $A^T A$ is invertible. If A does not have full rank, this means that the null space of A is not empty and $A^T A$ is not invertible. So, we need to use pseudoinverse. Furthermore, since the null space of A is not empty, there exist $v \in (A)$ such that $A(x + v) = Ax$. Therefore, we can add elements of (A) into the solution and still satisfies the equation $A^T Ax = A^T b$. Thus, the solutions are $x = (A^T A)^{-1} A^T b + v$, where $v \in \text{null}(A)$.

Number 5

A function f is β-smooth when its gradient is β-Lipschitz continuous.

(a) Find a global bound for β of the least-squares objective $\frac{1}{2}\|Ax - b\|^2$.

(b) Find a global bound for β of the regularized logistic objective

$$\sum_{i=1}^{n} \log(1 + \exp(\langle a_i, x\rangle)) + \frac{\lambda}{2}\|x\|^2.$$

(c) Do the gradients for Poisson regression admit a global Lipschitz constant?

1. By following the methods described in class, we find the β in $\|\nabla f(x) - \nabla f(y)\| \leq \beta\|(x - y)\|$ where f is the objective function.

$$\begin{aligned}
\|\nabla f(x) - \nabla f(y)\| &= \|A^T Ax - A^T Ay\| \\
&= \|A^T A(x - y)\| \\
&\leq \|A^T A\|\|x - y\| \\
&= \lambda_{\max}(A^T A)\|x - y\| \quad \text{(2-norm definition)}
\end{aligned}$$

So, $\beta = \lambda_{\max}(A^T A)$.

2. We can calculate the hessian of the objective function and find the bound of the norm of the hessian.

$$\nabla^2 f(x) = \sum_{i=1}^{n} \frac{\exp(a_i^T x)}{(1 + \exp(a_i^T x))^2} a_i a_i^T + \lambda \mathbf{I} \quad \text{(from 1(b))} \tag{1}$$

The term $\frac{\exp(a_i^T x)}{(1+\exp(a_i^T x))^2} \leq 0.25$ based on the plot of $\frac{\exp(y)}{(1+\exp(y))^2}$ where y is the scalar value based on the dot product result.

$$\begin{aligned}
\nabla^2 f(x) &= \sum_{i=1}^{n} \frac{\exp(a_i^T x)}{(1 + \exp(a_i^T x))^2} a_i a_i^T + \lambda \mathbf{I} \\
&\leq 0.25 \sum_{i=1}^{n} a_i a_i^T + \lambda \mathbf{I} \\
&= 0.25 A^T A + \lambda \mathbf{I}
\end{aligned}$$

Now, we can find the bound for the norm.

$$\begin{aligned}\|\nabla^2 f(x)\| &= \|0.25A^TA + \lambda\mathbf{I}\| \\ &\leq 0.25\|A^TA\| + \lambda\|\mathbf{I}\| \\ &= 0.25\lambda_{\max}(A^TA) + C\end{aligned}$$

So, $\beta = 0.25\lambda_{\max}(A^TA) + C$, where C is the regularized parameter.

3. We can calculate the hessian of the Poisson regression

$$\nabla^2 \left[\sum_{i=1}^{n} \exp(a_i^T x) - b^T Ax\right] = \sum_{i=1}^{n} a_i \exp(a_i^T x) a_i^T$$

From the expression we can see that the function exp is unbounded so when we find the norm, there will be no upper bound, so there will be no global Lipschitz constant.

Number 6

Please complete the coding homework (starting with the notebook uploaded to Canvas).