Recitation 12

https://cims.nyu.edu/ cd2754/

Fall 2021

Convex functions inequality

Proof the following inequality:
Given convex function f defined on $\mathbb{R}^n$, show for any x, y in the domain of f

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

$\forall\ x, y \in \mathbb{R}^n$, $\alpha \in (0,1)$ we have.

$$f(\alpha y + (1-\alpha)x) \leq \alpha f(y) + (1-\alpha) f(x)$$

Then.

$$f(x + \alpha(y-x)) \leq f(x) + \alpha(f(y) - f(x))$$

$$f(y) \geqslant f(x) + \frac{f(x+\alpha(y-x)) - f(x)}{\alpha}.$$

as $\alpha \to 0$, we have.

$$f(y) \geqslant f(x) + f(x)^T (y-x)$$

Convex functions and convex sets

For $f : \mathbb{R}^n \to \mathbb{R}$, define the epigraph $\text{epi}(f) \subset \mathbb{R}^{n+1}$ to be the set of all points above the graph of f:
$\text{epi}(f) := \{(x, t) \in \mathbb{R}^{n+1} \mid t \geq f(x)\}$.

1. Prove that f is convex if and only if $\text{epi}(f)$ is convex.
2. Prove that if f, g are convex functions, then $h(x) = \max(f(x), g(x))$ is convex.

1. Let $x_1, x_2 \in \mathbb{R}^n$, $y_1, y_2 \in \mathbb{R}$ $\alpha \in (0,1)$ and (x_1, y_1) (x_2, y_2) $\in \text{epi}(f)$
i.e. $y_1 \geq f(x_1)$ $y_2 \geq f(x_2)$

Then we have.

$\Rightarrow$ $\alpha y_1 + (1-\alpha) y_2 \geq \alpha f(x_1) + (1-\alpha) f(x_2) \geq f(\alpha x_1 + (1-\alpha) x_2)$

$(x_1, y_1), (x_2, y_2) \in \text{epi}(f) \Rightarrow (\alpha x_1 + (1-\alpha) x_2, \alpha y_1 + (1-\alpha) y_2) \in \text{epi}(f)$, $\text{epi}(f)$ convex.

$\Leftarrow$ since $(x_1, f(x_1))$; $(x_2, f(x_2)) \in \text{epi}(f)$, then $(\alpha x_1 + (1-\alpha) x_2, \alpha f(x_1) + (1-\alpha) f(x_2)) \in \text{epi}(f)$

$\alpha f(x_1) + (1-\alpha) f(x_2) \geq f(\alpha x_1 + (1-\alpha) x_2)$

f convex.

2. f, g convex

$\Rightarrow \mathrm{epi}(f), \mathrm{epi}(g)$ convex

Intersection of convex sets is convex

Then. $\mathrm{epi}(\max(f, g)) = \mathrm{epi}(f) \cap \mathrm{epi}(g)$ is convex

$\mathrm{epi}(h)$ convex $\Rightarrow h$ convex

Questions: Ridge Regression

Let $X \in \mathbb{R}^{n \times d}$, $n > d$, and *not have full rank*. (X is a data matrix)
Recall that the OLS solution is $\hat{x} = (X^T X)^{-1} X^T y$.

1. Since X is not full rank, what does this say about the features?
2. What is the issue with the OLS solution?
3. The ridge regression solution is given by $(X^T X + \lambda Id_d)^{-1} X^T y$. How does this fix the issue?
4. Suppose that X has SVD $X = U \Sigma V^T$, and X has singular values $\sigma_1, \ldots, \sigma_d$. What are the eigenvalues of $X^T X + \lambda Id_d$?
5. How does increasing λ affect the condition number of $(X^T X + \lambda Id_d)$?

1. linear dependent columns, duplicated features.

2. Non unique solution, problem is not well-defined.

3. $(X^T X + \lambda I_d)^{-1} X^T y$ ensures a unique solution as $(X^T X + \lambda I_d)$ is invertible.

4. Suppose $\sigma_1 \dots \sigma_d$ are singular values of X, then

eigenvalue of $X^T X$: $\sigma_1^2 \dots \sigma_d^2$

eigenvalue of $(X^T X + \lambda I_d)$: $\sigma_1^2 + \lambda, \dots \sigma_d^2 + \lambda$

Note σ_i may be 0.

5. condition number : $\frac{\max \lambda_i}{\min \lambda_i}$

The higher condition number is, the more "unstable" the matrix is.

The transformation is much higher along some dimension than others.

Creates problem when solving/optimizing.

is X has zero-valued singular value.

Condition number of $X^T X = \frac{\sigma_1^2}{\sigma_d^2} = \frac{c}{0} = \infty$ (very unstable)

Condition number of $X^T X + \lambda I_d = \frac{\sigma_1^2 + \lambda}{\sigma_d^2 + \lambda} < \infty$

We can adjust the condition number by changing λ.

as $\lambda \to \infty$, $\frac{\sigma_1^2 + \lambda}{\sigma_d^2 + \lambda} \to 1$