RECITATION 3: LINEAR REGRESSION & OPTIMIZATION

10-301/10-601 Introduction to Machine Learning (Summer 2023)
http://www.cs.cmu.edu/~hchai2/courses/10601
Released: June 1st, 2023
Quiz Date: June 6th, 2023
TAs: Alex, Andrew, Sofia, Tara, and Neural the Narwhal

1 Quiz Preparation

1.1 Objective Functions

1. In the context of linear regression, what does an objective function $\ell(\mathbf{w})$ do?

Your answer:

2. What are some desirable properties of a good objective function?

Your answer:

1.2 Closed-form Solution for Linear Regression

Suppose we are given the following dataset where x is the input and y is the output:

x	1.0	2.0	3.0	4.0	5.0
y	2.0	4.0	7.0	8.0	11.0

Based on our inductive bias, we think that the linear hypothesis with no intercept should be used here. We also want to use the Mean Squared Error as our objective function: $\frac{1}{5}\sum_{i=1}^{5}(y^{(i)} - wx^{(i)})^2$, where $y^{(i)}$ is our i^{th} data point and w is our weight. Using the closed-form method, find w.

1. What is the closed-form formula for w?

Your answer:

2. What is the value of w?

Your answer:

Now let's extend the data set to include more features, $\mathbf{x} \in \mathbb{R}$:

	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	$\mathbf{x}^{(3)}$	$\mathbf{x}^{(4)}$	$\mathbf{x}^{(5)}$
x_1	1.0	2.0	3.0	4.0	5.0
x_2	-2.0	-5.0	-6.0	-8.0	-11.0
x_3	3.0	8.0	9.0	12.0	14.0
y	2.0	4.0	7.0	8.0	11.0

We again think that a linear hypothesis with no bias should be used here. We also want to use the Mean Squared Error as our objective function:

$$\frac{1}{N}\sum_{i=1}^{N}(y^{(i)} - \mathbf{w}^T\mathbf{x}^{(i)})^2,$$

where $\mathbf{w} = [w_1, w_2, w_3]^T$, $\mathbf{x}^{(i)}$ is the i^{th} datapoint and $y^{(i)}$ is the i^{th} $y-$value.

1. What are the design matrix X and target vector $\mathbf{y}$ in this setting?

Your answer:

2. What is the closed-form matrix solution for $\mathbf{w}$?

Your answer:

1.3 Gradient Descent for Linear Regression

Consider the following dataset:

	$\mathbf{x}^{(1)}$	$\mathbf{x}^{(2)}$	$\mathbf{x}^{(3)}$	$\mathbf{x}^{(4)}$	$\mathbf{x}^{(5)}$
x_1	1.0	2.0	3.0	4.0	5.0
x_2	-2.0	-5.0	-6.0	-8.0	-11.0
y	2.0	4.0	7.0	8.0	11.0

Suppose we want to implement gradient descent using a stepsize of $\eta = 0.1$. Assuming $\mathbf{w}$ has been initialized to $[0, 0, 0]^T$, let's perform one iteration of gradient descent:

1. What is the gradient of the objective function $\ell(\mathbf{w})$ with respect to $\mathbf{w}$: $\nabla_{\mathbf{w}} \ell(\mathbf{w})$?

Your answer:

2. How do we carry out the update rule?

Your answer:

3. How could we pick which value of η to use if we weren't given the step size?

Your answer: