
DSC 140A - Quiz 02

Name:

PID:

Quiz 02 Instructions. All technology is prohibited. Pen and paper only. You may not discuss with any classmates. Any prohibited use of notes or technology will result in a 0 and is a violation of Academic Integrity. You will have 35 minutes to complete your quiz. The rest of the discussion time, we can go over the quiz together. If you have any questions, please raise your hand and a TA will come assist you. Make sure to write your name and PID at the top of every page as clearly as possible. Explain/show your work.

Notation: Bold upper case letters \mathbf{A} represent matrices.
Bold lower case letters \mathbf{a} represent vectors.
Non-bolded lower case letters a represent scalars.
For a matrix \mathbf{A} , \mathbf{a}_i indicates the i th row vector of matrix \mathbf{A} .

Problem 1.

Convexity

- a) Prove from the definition of convexity that $f(\mathbf{x}) = \|\mathbf{x}\|_2^2$ is convex on \mathbb{R}^n . Do not use the second derivative test.

Solution:

We must show that for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and all $\lambda \in [0, 1]$:

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}).$$

Step 1: Expand the left-hand side.

Recall that $\|\mathbf{z}\|_2^2 = \mathbf{z}^\top \mathbf{z}$, so:

$$\begin{aligned} f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) &= \|\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}\|_2^2 \\ &= (\lambda\mathbf{x} + (1 - \lambda)\mathbf{y})^\top (\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \\ &= \lambda^2 \|\mathbf{x}\|_2^2 + 2\lambda(1 - \lambda) \mathbf{x}^\top \mathbf{y} + (1 - \lambda)^2 \|\mathbf{y}\|_2^2. \end{aligned}$$

Step 2: Expand the right-hand side.

$$\lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}) = \lambda \|\mathbf{x}\|_2^2 + (1 - \lambda) \|\mathbf{y}\|_2^2.$$

Step 3: Compute the difference RHS – LHS.

$$\begin{aligned} &\lambda \|\mathbf{x}\|_2^2 + (1 - \lambda) \|\mathbf{y}\|_2^2 - \lambda^2 \|\mathbf{x}\|_2^2 - 2\lambda(1 - \lambda) \mathbf{x}^\top \mathbf{y} - (1 - \lambda)^2 \|\mathbf{y}\|_2^2 \\ &= (\lambda - \lambda^2) \|\mathbf{x}\|_2^2 - 2\lambda(1 - \lambda) \mathbf{x}^\top \mathbf{y} + ((1 - \lambda) - (1 - \lambda)^2) \|\mathbf{y}\|_2^2 \\ &= \lambda(1 - \lambda) \|\mathbf{x}\|_2^2 - 2\lambda(1 - \lambda) \mathbf{x}^\top \mathbf{y} + \lambda(1 - \lambda) \|\mathbf{y}\|_2^2. \end{aligned}$$

Step 4: Factor.

$$\begin{aligned} &= \lambda(1 - \lambda) (\|\mathbf{x}\|_2^2 - 2 \mathbf{x}^\top \mathbf{y} + \|\mathbf{y}\|_2^2) \\ &= \lambda(1 - \lambda) \|\mathbf{x} - \mathbf{y}\|_2^2. \end{aligned}$$

Step 5: Conclude.

Since $\lambda \in [0, 1]$, we have $\lambda \geq 0$ and $(1 - \lambda) \geq 0$. Also, $\|\mathbf{x} - \mathbf{y}\|_2^2 \geq 0$ always. Therefore:

$$\text{RHS} - \text{LHS} = \lambda(1 - \lambda) \|\mathbf{x} - \mathbf{y}\|_2^2 \geq 0.$$

This gives us:

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}),$$

and so $f(\mathbf{x}) = \|\mathbf{x}\|_2^2$ is convex on \mathbb{R}^n .

- b) Does convexity guarantee that (sub)gradient descent will reach a global minimum? Explain.

Solution: Yes. Convexity guarantees that every local minimum is a global minimum, and even if the function is non-differentiable, sub-gradient descent will still allow us to find the minima.

Problem 2.

Fill in all correct descriptions of stochastic gradient descent. Let N be the number of total training samples and n be the number of samples the gradient is calculated over.

- $n = N$
- $n < N$
- Perform Gradient descent on first n samples, then next n samples
- Randomly select n samples every gradient descent step

Problem 3.

Given:

- One training example: $\mathbf{x} = [1, 2]^T$ with true label $y = 10$
- A linear model with no intercept: $f_{\mathbf{w}}(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$
- Initial weights: $w_1 = 1, w_2 = 2$
- Learning rate: $\alpha = 0.1$
- MSE Loss function: $L(\mathbf{w}, \mathbf{x}) = \frac{1}{n} \sum_{i \in [n]} (f_{\mathbf{w}}(\mathbf{x}) - y)^2$

a) Compute the model's prediction \hat{y} using the initial weights.

Solution: The prediction is given by:

$$\hat{y} = f_{\mathbf{w}}(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$$

Substitute $\mathbf{w} = [1, 2]^T$ and $\mathbf{x} = [1, 2]^T$:

$$\hat{y} = w_1 x_1 + w_2 x_2 = (1)(1) + (2)(2) = 1 + 4 = 5$$

$$\boxed{\hat{y} = 5}$$

b) Compute the loss L .

Solution:

Since we have only $n = 1$ training example, the MSE loss simplifies to:

$$L(\mathbf{w}) = \frac{1}{1} (\hat{y} - y)^2 = (\hat{y} - y)^2$$

Substitute $\hat{y} = 5$ and $y = 10$:

$$L(\mathbf{w}) = (5 - 10)^2 = (-5)^2 = 25$$

$$\boxed{L = 25}$$

c) Compute the gradient of the loss.

Solution:

Write the loss in terms of the weights:

$$L(\mathbf{w}) = (\mathbf{w}^\top \mathbf{x} - y)^2$$

Apply the chain rule. Let $e = \mathbf{w}^\top \mathbf{x} - y$ denote the error, so $L = e^2$:

$$\frac{\partial L}{\partial w_j} = 2e \cdot \frac{\partial e}{\partial w_j}$$

Since $e = w_1x_1 + w_2x_2 - y$, we have $\frac{\partial e}{\partial w_j} = x_j$. Therefore:

$$\frac{\partial L}{\partial w_j} = 2(\mathbf{w}^\top \mathbf{x} - y) x_j$$

In vector form:

$$\nabla_{\mathbf{w}} L = 2(\mathbf{w}^\top \mathbf{x} - y) \mathbf{x}$$

Substitute $e = \hat{y} - y = 5 - 10 = -5$ and $\mathbf{x} = [1, 2]^\top$:

$$\nabla_{\mathbf{w}} L = 2(-5) \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -10 \\ -20 \end{bmatrix}$$

$$\boxed{\nabla_{\mathbf{w}} L = \begin{bmatrix} -10 \\ -20 \end{bmatrix}}$$

d) Update the weights using one step of gradient descent.

Solution:

The gradient descent update rule is:

$$\mathbf{w}^{\text{new}} = \mathbf{w}^{\text{old}} - \alpha \nabla_{\mathbf{w}} L$$

Substitute $\mathbf{w}^{\text{old}} = [1, 2]^\top$, $\alpha = 0.1$, and $\nabla_{\mathbf{w}} L = [-10, -20]^\top$:

$$\mathbf{w}^{\text{new}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - 0.1 \begin{bmatrix} -10 \\ -20 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

The updated weights are:

$$\boxed{w_1^{\text{new}} = 2, \quad w_2^{\text{new}} = 4}$$