Quiz-2: study guide

This quiz is largely a DSAN-5000 review, however the content supports upcoming DSAN-5300 material.

Problem-1: Two-Variable Optimization

1. Problem Setup

You are given a scalar objective (loss) function of two variables:

$L(\mathbf{w}) = L(w_0, w_1)$

Your goal is to find values of (w_0) and (w_1) that optimize (minimize or maximize) (L).

---

2. Necessary Condition for an Optimum

A **necessary condition** for a local minimum, maximum, or stationary point is that the gradient is zero:

$\nabla L(w_0, w_1) = \mathbf{0}$

This does **not** guarantee a minimum or maximum—it identifies *critical points*.

---

3. Computing the Gradient

The gradient of (L(w_0, w_1)) is the vector of [1] partial derivatives:

$$\nabla L = \left[\boxed{\frac{\partial L}{\partial w_0}} \; \frac{\partial L}{\partial w_1} \right] \tag{1}$$

1. This is partial derivative. messures how L changes as w0 changes while w1 is held fixed

---

4. Solving for Critical Points

1. Compute $(\frac{\partial L}{\partial w_0}) and (\frac{\partial L}{\partial w_1})$
2. Set each partial derivative equal to zero
3. Solve the resulting system of equations for (w_0, w_1), which corresponds to the location of the critical point w_0^*, w_1^*

---

5. Evaluating the Objective at the Optimum

Once found, compute: $L(w_0^*, w_1^*)$, This is the *value of the objective function at the optimum*.

Worked Example

Given Objective Function

$$L\left(w_0, w_1\right) = w_0^2 + 2w_1^2 - 4w_0 - 8w_1 + 10 \tag{2}$$

Step 1: Compute the Gradient

$$\begin{aligned} \frac{\partial L}{\partial w_0} &= 2w_0 - 4 \\ \frac{\partial L}{\partial w_1} &= 4w_1 - 8 \end{aligned} \tag{3}$$

This is how to calculate partial derivatives for w0

Step 2: Set Gradient Equal to Zero

$$\begin{aligned} 2w_0 - 4 = 0 \\ 4w_1 - 8 = 0 \end{aligned} \tag{4}$$

Step 3: Solve the System

$$\begin{aligned} w_0 = 2 \\ w_1 = 2 \end{aligned} \tag{5}$$

Step 4: Evaluate the Objective at the Optimum

$$\begin{aligned} L(2,2) &= (2)^2 + 2(2)^2 - 4(2) - 8(2) + 10 \\ &= 4 + 8 - 8 - 16 + 10 \\ &= -2 \end{aligned} \tag{6}$$

Final Answer

- Optimal parameters: $w_0 = 2, w_1 = 2$
- Objective value at optimum: $L = -2$

Problem-2: Single parameter fitting

1. Model Specification

You are given a linear model with one fitting parameter:

$$\hat{y}_i = wx_i \tag{7}$$

where:

- x_i is a known input
- y_i is the observed target

- w is the parameter to be learned

2. Training Data

The training set consists of n input-output pairs:

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \tag{8}$$

3. Define the Loss Function (MSE)

The mean squared error (MSE) loss is:

$$L(w) = \frac{1}{n} \sum_{i=1}^{n} (y_i - wx_i)^2 \tag{9}$$

This is a scalar function of a single variable w.

4. Standard Regression Workflow
5. Model: $\hat{y}_i = wx_i$
6. Loss: $L(w) = \frac{1}{n} \sum (y_i - wx_i)^2$
7. Optimize: Differentiate $L(w)$ with respect to w
8. Estimate: Solve for w by setting the derivative equal to zero
9. Compute the Derivative of the Loss

Differentiate $L(w)$ with respect to w :

$$\frac{dL}{dw} = \frac{1}{n} \sum_{i=1}^{n} 2\,(y_i - wx_i)\,(-x_i) \tag{10}$$

Simplify:

$$\frac{dL}{dw} = -\frac{2}{n} \sum_{i=1}^{n} x_i\,(y_i - wx_i) \tag{11}$$

6. Set Derivative Equal to Zero

$$-\frac{2}{n} \sum_{i=1}^{n} x_i\,(y_i - wx_i) = 0 \tag{12}$$

Drop the constant factor:

$$\sum_{i=1}^{n} x_i y_i - w \sum_{i=1}^{n} x_i^2 = 0 \tag{13}$$

7. Solve for the Optimal w

$$w^* = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \tag{14}$$

This is the closed-form solution for single-parameter linear regression through the origin.

Problem-3: Numerical implementation

Consider Training data

$$(x, y) = \{(1, 2), (2, 4), (3, 6)\}, \quad n = 3 \tag{15}$$

Model:

$$\hat{y} = wx \tag{16}$$

this is the model that need to be plotted in problem3 4.1 with known w

MSE loss:

$$L(w) = \frac{1}{n} \sum_{i=1}^{n} (y_i - wx_i)^2 \tag{17}$$

1. **Optimal w^* (pen-and-paper recap)**

Closed-form solution (through the origin):

$$w^* = \frac{\sum x_i y_i}{\sum x_i^2} \tag{18}$$

Compute:

$$\sum x_i y_i = 1 \cdot 2 + 2 \cdot 4 + 3 \cdot 6 = 28, \quad \sum x_i^2 = 1^2 + 2^2 + 3^2 = 14 \tag{19}$$

So:

$$w^* = \frac{28}{14} = 2 \tag{20}$$

2. **Write the loss curve $L(w)$ explicitly (parameter space)**

Plug the data into the MSE expression:

$$L(w) = \frac{1}{3}\left[(2 - w)^2 + (4 - 2w)^2 + (6 - 3w)^2\right] \tag{21}$$

Expand/simplify:

$$L(w) = \frac{14}{3}w^2 - \frac{56}{3}w + \frac{56}{3} = \frac{14}{3}(w - 2)^2 \tag{22}$$

Minimum (in parameter space)
From $\frac{14}{3}(w-2)^2$, the minimum is clearly at:

$$w^* = 2, \quad L(w^*) = 0 \tag{23}$$

3. **Compute MSE manually at the optimum (data space MSE from residuals)**

With $w^* = 2$, predictions:

$$\hat{y}_1 = 2 \cdot 1 = 2, \quad \hat{y}_2 = 2 \cdot 2 = 4, \quad \hat{y}_3 = 2 \cdot 3 = 6 \tag{24}$$

Residuals:

$$y_i - \hat{y}_i = 0, 0, 0 \tag{25}$$

Squared residuals sum:

$$0^2 + 0^2 + 0^2 = 0 \tag{26}$$

So the MSE is:

$$\mathrm{MSE} = \frac{1}{3} \cdot 0 = 0 \tag{27}$$

This matches the parameter-space result $L\left(w^*\right) = 0$, as it should

4. Plotting:
    1. Plot in data-space: x-y plot with data points, and optimal model
    2. Plot in parameter space: w-L(w), plot loss curve with numeric values, label location of minima, and value of MSE on axis