

SMAI-M20-Lec 13 Review questions

IIIT Hyderabad

September 9, 2020

Review Question - I (one, none or more correct)

- ① (use notations and conventions from the class) Consider the problem of linear regression where we minimize the loss

$$\mathcal{L}_1 = \frac{1}{N} \sum_{i=1}^N \alpha_i (y_i - \mathbf{w}^T \mathbf{x}_i)^2 + \lambda_1 g(\mathbf{w})$$

where $g()$ is a regularization term. We also write the loss in matrix form as

$$\mathcal{L}_2 = \frac{1}{N} [\mathbf{Y} - \mathbf{X}\mathbf{w}]^T \mathbf{A} [\mathbf{Y} - \mathbf{X}\mathbf{w}] + \lambda_2 g(\mathbf{w}).$$

If $\mathcal{L}_1 = \mathcal{L}_2$ for all \mathbf{w} , then

- ① \mathbf{A} is a diagonal matrix
- ② $A_{ij} = \alpha_i \cdot \alpha_j$
- ③ $A_{ii} = \alpha_i$ else zero
- ④ $A_{ii} = \frac{1}{\alpha_i}$ else zero
- ⑤ none of the above

Ans: AC

Review Question - II (one, none or more correct)

(use notations and conventions from the class) Consider the problem of linear regression where we minimize the loss

$$\mathcal{L}_1 = \frac{1}{N} \sum_{i=1}^N \alpha_i (y_i - \mathbf{w}^T \mathbf{x}_i)^2 + \lambda_1 g(\mathbf{w})$$

where $g()$ is a regularization term. We also write the loss in matrix form as

$$\mathcal{L}_2 = \frac{1}{N} [Y - \mathbf{X}\mathbf{w}]^T A [Y - \mathbf{X}\mathbf{w}] + \lambda_2 g(\mathbf{w}).$$

If $\mathbf{A} = I$, $\alpha_i = 1$ for all i , and $\lambda_1 = \lambda_2 = 1$, then

- ① Both the loss functions are identical i.e., $\mathcal{L}_1 = \mathcal{L}_2$
- ② The optima of the first objective \mathbf{w}_1^* is same as the optima of \mathcal{L}_2 , i.e., \mathbf{w}_2^*
- ③ At the optima, value of the losses are same. i.e., $\mathcal{L}_1^* = \mathcal{L}_2^*$
- ④ \mathcal{L}_1 is a scalar and \mathcal{L}_2 is a vector
- ⑤ none of the above

Ans: ABC

Review Question - III (one, none or more correct)

(use notations and conventions from the class) Consider the problem of linear regression where we minimize the loss

$$\mathcal{L}_1 = \frac{1}{N} \sum_{i=1}^N \alpha_i (y_i - \mathbf{w}^T \mathbf{x}_i)^2 + \lambda_1 g(\mathbf{w})$$

where $g()$ is a regularization term. We also write the loss in matrix form as

$$\mathcal{L}_2 = \frac{1}{N} [\mathbf{Y} - \mathbf{X}\mathbf{w}]^T \mathbf{A} [\mathbf{Y} - \mathbf{X}\mathbf{w}] + \lambda_2 g(\mathbf{w}).$$

If $\mathbf{A} = \mathbf{I}$, $\alpha_i = 2$ for all i , and $\lambda_1 = \lambda_2 = 0$, then

- ① Both the loss functions are identical i.e., $\mathcal{L}_1 = \mathcal{L}_2$
- ② The optima of the first objective \mathbf{w}_1^* is same as the optima of \mathcal{L}_2 , i.e., \mathbf{w}_2^*
- ③ At the optima, value of the losses are same. $\mathcal{L}_1^* = \mathcal{L}_2^*$
- ④ \mathcal{L}_1 is a scalar and \mathcal{L}_2 is a vector
- ⑤ none of the above

Ans: B $\mathcal{L}_1 = 2 * \mathcal{L}_2$ but the optimal weights remain same

Review Question - IV (one, none or more correct)

(use notations and conventions from the class) Consider the problem of linear regression where we minimize the loss

$$\mathcal{L}_1 = \frac{1}{N} \sum_{i=1}^N \alpha_i (y_i - \mathbf{w}^T \mathbf{x}_i)^2 + \lambda_1 g(\mathbf{w})$$

where $g()$ is a regularization term. We also write the loss in matrix form as

$$\mathcal{L}_2 = \frac{1}{N} [Y - \mathbf{X}\mathbf{w}]^T A [Y - \mathbf{X}\mathbf{w}] + \lambda_2 g(\mathbf{w}).$$

If $\mathbf{A} = I$, $\alpha_i = 1$ for all i , and $\lambda_1 \neq \lambda_2 \neq 0$, then

- ① The optimal parameters \mathbf{w}^* is independent of λ_i .
- ② The larger the lambda, the better the solution.
- ③ The smaller the lambda, the better the solution
- ④ When lambda is nonzero (positive), loss will increase (since $g(w)$ is also positive in practice), better to use $\lambda = 0$.
- ⑤ None of the above.

Ans: E

Review Question - V (one, none or more correct)

(use notations and conventions from the class) Consider the problem of linear regression where we minimize the loss

$$\mathcal{L}_1 = \frac{1}{N} \sum_{i=1}^N \alpha_i (y_i - \mathbf{w}^T \mathbf{x}_i)^2 + \lambda_1 g(\mathbf{w})$$

where $g()$ is a regularization term. We also write the loss in matrix form as

$$\mathcal{L}_2 = \frac{1}{N} [Y - \mathbf{X}\mathbf{w}]^T A [Y - \mathbf{X}\mathbf{w}] + \lambda_2 g(\mathbf{w}).$$

See \mathcal{L}_2 closely,

- 1 When A is a diagonal matrix, this is equivalent to weighing each sample independently.
- 2 When A is not a diagonal matrix, this loss does not make any sense. Don't use.
- 3 When A is PD, we can do cholesky decomposition of A as LL^T and an equivalent formulation is possible in \mathcal{L}_1 is each sample getting transformed as $\mathbf{L}^T \mathbf{x}_i$ (as in LMNN/Metric Learning)

- ④ When A is a rank deficient matrix, an equivalent formulation is possible in \mathcal{L}_1 with a dimensionality reduction (this could be proved with eigen decomposition).
- ⑤ None of the above

Ans: ACD