

CS583A: Quiz 1 (Sample Questions)

Name:

Policy: Books and printed materials are allowed. Do not use electronic device, including phone, laptop, and tablet.

Hint: (i) $\frac{\partial e^a}{\partial a} = e^a$, (ii) $\frac{\partial \log_e(a)}{\partial a} = \frac{1}{a}$, (iii) $\frac{\partial \frac{1}{a}}{\partial a} = -\frac{1}{a^2}$, and (iv) $\frac{\partial \cos(a)}{\partial a} = -\sin(a)$.

Q1 (12%). Let \mathbf{A} be the 3×3 diagonal matrix:

$$\mathbf{A} = \text{diag}([1, 2, 5]) \triangleq \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

Calculate the following values:

1. the trace:

$$\text{tr}(\mathbf{A}) =$$

2. the squared Frobenius-norm of \mathbf{A} :

$$\|\mathbf{A}\|_F^2 =$$

3. the condition number of \mathbf{A} :

$$\kappa(\mathbf{A}) =$$

4. the rank of \mathbf{A} :

$$\text{rank}(\mathbf{A}) =$$

Q2 (16%). Let $\sigma_1 = 5$, $\sigma_2 = 5$, $\sigma_3 = 3$, $\sigma_4 = 2$, and $\sigma_5 = 1$. Let $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_5\} \subset \mathbb{R}^{100}$ be an orthonormal basis of some subspace. Let $\mathbf{A} = \sum_{i=1}^5 \sigma_i \mathbf{v}_i \mathbf{v}_i^T \in \mathbb{R}^{100 \times 100}$. Calculate the following values:

1. the squared Frobenius norm of \mathbf{A} :

$$\|\mathbf{A}\|_F^2 =$$

2. the matrix-vector product:

$$\mathbf{A} \mathbf{v}_3 =$$

3. matrix rank:

$$\text{rank}(\mathbf{A}) =$$

4. Let $\mathbf{B} = \text{argmin}_{\text{rank}(\mathbf{X}) \leq 2} \|\mathbf{A} - \mathbf{X}\|_F^2$. Calculate the squared Frobenius norm distance:

$$\|\mathbf{A} - \mathbf{B}\|_F^2 =$$

Q3 (3%). Let $f : \mathbb{R}^d \mapsto \mathbb{R}$ be a convex function. A local minimum of $\min_{\mathbf{w}} f(\mathbf{w})$ is also a global minimum.

- A. The statement is true.
- B. The statement is false.

Q4 (3%). The set $\{\mathbf{x} \in \mathbb{R}^{100} \mid \|\mathbf{x}\|_1 = 5\}$ is a convex set.

- A. The statement is true.
- B. The statement is false.

Q5 (3%). As we increase the model capacity (e.g., the number of layers of a convolutional neural network), the training error and test error both increase. What is happening when we increase the model capacity?

- A. Overfitting.
- B. Underfitting.
- C. Neither of the above.

Q6 (3%). Consider the following ridge regression model:

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \frac{\gamma}{2} \|\mathbf{w}\|_2^2,$$

where $\gamma \geq 0$. The Hessian matrix is $\mathbf{H} = \mathbf{X}^T \mathbf{X} + \gamma \mathbf{I}_d$. As γ grows, what will happen to $\kappa(\mathbf{H})$ (the condition number of \mathbf{H}).

- A. The condition number increases.
- B. The condition number decreases.
- C. Both of the above are possible.
- D. Neither of the above are possible.

(Hint: The condition number of \mathbf{H} is $\frac{\sigma_{\max}(\mathbf{X}^T \mathbf{X}) + \gamma}{\sigma_{\min}(\mathbf{X}^T \mathbf{X}) + \gamma} \geq 1$.)

Q7 (3%). Trained using the same samples and using the same computation time and memory, classification model \mathcal{M}_1 achieves a test accuracy 98%, while classification model \mathcal{M}_2 achieves a test accuracy 70%. Model \mathcal{M}_1 is very likely a better choice than model \mathcal{M}_2 .

- A. The statement is true.
- B. The statement is false.

Q8 (3%). What is the subdifferential of the function $f(w) = |w|$ at $w = 0$?

- A. -1 .
- B. 0 .
- C. 1 .
- D. $\{0\}$.
- E. $[-1, 1]$.

Detach this empty page and use it for calculation. Do NOT submit this page.

Detach this empty page and use it for calculation. Do NOT submit this page.