1 Multiple Choice 1 point

What does "linear" in "linear regression" mean?

- [ ] The computational scaling of fitting a model is linear in the number of examples
- [x] The relationship between dependent variable and independent variables is assumed to be linear
- [ ] The loss function is a linear function
- [ ] The number of examples scales linearly with the number of features

2 Multiple Choice 1 point

If there are n examples and d features in a linear regression problem, then what are the dimensions of the design matrix? For this problem, assume that we have NOT aded a dummy feature to absorb the bias into the weight vector.

- [x] n x d
- [ ] d x n
- [ ] n x 1
- [ ] d x 1

$$\begin{array}{cc} & d \\ n & \begin{bmatrix} x_1^{(1)} & \cdots & x_d^{(1)} \\ \vdots & & \\ x_1^{(n)} & \cdots & x_d^{(n)} \end{bmatrix} \end{array} + \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

3 Multiple Choice 1 point

Which of the following is NOT a good loss function? Recall our convention that y_hat is the prediction and y is the true value and a good loss function should penalize predictions that are far from the true value.

- [x] $(y\_hat)^2 + y^2 + 2\,(y\_hat)\,y$ X
- [ ] $(y\_hat)^2 + y^2 - 2\,(y\_hat)\,y$
- [ ] $(y\_hat - y)^2$
- [ ] |y_hat - y| where |x| = max{x, -x} denotes the absolute value of x

4 Multiple Choice 1 point

How are models trained in deep learning?

- [x] By minimizing a loss over the choice of model parameters
- [ ] By maximizing a loss over the choice of model parameters
- [ ] By randomly choosing the values of model parameters
- [ ] By having an expert engineer manually choose model parameters

5 Matching 1 point

There are four important objects in a linear regression set-up. The design matrix **X**, the vector of true labels **y**, the vector of model predictions **y-hat**, and the model parameters **w**. Assume that bias has been absorbed into **w** by adding a dummy "always one" feature in the design matrix and that the feature dimension after doing this is d. Also assume that the number of examples is n. Match the objects to their dimensions

design matrix	n x d
vector of true labels	n x 1
vector of model predictions	n x 1
vector of model parameters	d x 1

Possible answers

- n x 1
- n x d
- d x 1

6 Multiple Choice 1 point

The closed-form formula for the minimizer $\mathbf{w}^*$ of $1/2 \, || \mathbf{y} - \mathbf{X} \mathbf{w} ||^2$ is $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$. How does this formula change if we change the objective function from $1/2 \, || \mathbf{y} - \mathbf{X} \mathbf{w} ||^2$ to $1/2 \, || \mathbf{y} - \mathbf{X} \mathbf{w} ||^2 + 1/2$?

- ○ $(\mathbf{X}^T\mathbf{X} + \mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}$ where **I** is the d x d identity matrix
- ● The formula doesn't change   + const 不影响 gradients
- ○ There is no longer any closed-form formula
- ○ $(\mathbf{X}^T\mathbf{X} + \mathbf{h})^{-1}\mathbf{X}^T\mathbf{y}$ where **h** is the d x d matrix with all entries equal to 1/2

7 Multiple Choice 1 point

The minimizer of the loss in linear regression has a closed form formula when we use the squared error (y-hat - y)^2 as our loss function. Suppose we switch to the absolute error |y-hat - y| as our loss function. That is, our loss function just measures the magnitude of the error without paying attention to the sign of the error. How does the closed form formula change as a result of changing the loss function from squared error to absolute error?

- There is no longer any closed-form formula
- The formula doesn't change
- $(\mathbf{X}^T\mathbf{X} + \mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}$ where **I** is the d x d identity matrix
- $(\mathbf{X}^T\mathbf{X} + \mathbf{1})^{-1}\mathbf{X}^T\mathbf{y}$ where **1** is the d x d matrix with all entries equal to 1

因为 $|\hat{y}-y|$ 在0点不是 differentiable 的

8 Multiple Choice 1 point

What is the gradient (i.e., vector of partial derivatives w.r.t. components of the input vector **w**) of 1/2 $\mathbf{w}^T\mathbf{w}$?

- 2 **w**
- **w**
- $\mathbf{w} + \mathbf{w}^T$
- $\mathbf{w}^T$

9 Multiple Choice 1 point

What is the gradient (i.e., vector of partial derivatives w.r.t. components of the input vector **w**) of $\mathbf{c}^T\mathbf{w}$?

- (selected) **c**
- **w**
- $\mathbf{c}^T\mathbf{w}$
- $\mathbf{c} + \mathbf{c}^T$

$\nabla_w c^T w = c$

10 Multiple Choice 1 point

$X^T X$ 一定 symmetric 且 square

Suppose **X** is an n x d design matrix. Which of the following is NOT always true about $\mathbf{X}^T\mathbf{X}$?

- It is square
- (selected) It is invertible
- It is symmetric
- It consists of inner products between feature vectors, i.e., column vector of the design matrix