Question 1

Not yet answered

Marked out of 1.00

Consider the predictor $f(x) = xw$, where $w \in \mathbb{R}$ is a one-dimensional parameter, and x represents the feature with no bias term. Suppose you are given a dataset of n data points $\mathcal{D} = ((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n))$, where each y_i is the target variable corresponding to feature x_i.

Let the loss function be the scaled squared loss $\ell(f(x), y) = c(f(x) - y)^2$ where $c \in \mathbb{R}$.

The estimate of the expected loss for a parameter $w \in \mathbb{R}$ is defined as the following convex function:

$$\hat{L}(w) = \frac{1}{n} \sum_{i=1}^{n} c(x_i w - y_i)^2$$

What is the closed form solution for $\hat{w} = \arg\min_{w \in \mathbb{R}} \hat{L}(w)$?

Select all that apply:

- [x] a. $\hat{w} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$ cross out
- [ ] b. $\hat{w} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$ cross out
- [ ] c. $\hat{w} = \frac{\sum_{i=1}^{n} y_i}{n}$ cross out
- [ ] d. $\hat{w} = \frac{\sum_{i=1}^{n} c x_i y_i}{\sum_{i=1}^{n} x_i^2}$ cross out

Question **2**

Not yet answered

Marked out of 1.00

Let everything be defined as in the previous question.
Suppose we consider the multivariate case where $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{w}$, and $\mathbf{w} \in \mathbb{R}^{d+1}$.
What is the closed form solution for $\hat{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathbb{R}^{d+1}} \hat{L}(\mathbf{w})$?

Select all that apply:

- [x] a. $\hat{\mathbf{w}} = A^{-1}b$ where $A = \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top$ and $b = \sum_{i=1}^n \mathbf{x}_i y_i$ (assume that A is invertible). cross out
- [ ] b. $\hat{\mathbf{w}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$ cross out
- [ ] c. $\hat{\mathbf{w}} = Ax$ where $A = \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top$ cross out
- [ ] d. $\hat{\mathbf{w}} = \frac{\sum_{i=1}^n c\mathbf{x}_i y_i}{\sum_{i=1}^n c\mathbf{x}_i^2}$ cross out

Question **3**

Not yet answered

Marked out of 1.00

Let $g(w) = -\ln w \sum_{i=1}^n y_i - \ln(1-w)\sum_{i=1}^n (1-y_i)$ where $w \in \mathbb{R}$.
We can rewrite this a bit more simply as $g(w) = -s\ln w - (n-s)\ln(1-w)$ where $s = \sum_{i=1}^n y_i$.
What is the derivative $g'(w)$ and the first order gradient descent update rule with a constant step size η?

Select all that apply:

- [ ] a. $g'(w) = -\frac{s}{1-w} + \frac{n-s}{w}$ and update rule $w \leftarrow w - \eta\left(-\frac{s}{1-w} + \frac{n-s}{w}\right)$ cross out
- [ ] b. $g'(w) = -\frac{s}{w} + \frac{n-s}{1-w}$ and update rule $w \leftarrow w - \eta\left(-\frac{s}{1-w} + \frac{n-s}{w}\right)$ cross out
- [ ] c. $g'(w) = -\frac{s}{1-w} - \frac{n-s}{w}$ and update rule $w \leftarrow w - \eta\left(-\frac{s}{1-w} - \frac{n-s}{w}\right)$ cross out
- [x] d. $g'(w) = -\frac{s}{w} + \frac{n-s}{1-w}$ and update rule $w \leftarrow w - \eta\left(-\frac{s}{w} + \frac{n-s}{1-w}\right)$ cross out

Question 4

Not yet answered

Marked out of 1.00

Let everything be defined as in the previous question.

What is the second derivative $g''(w)$ and the second order gradient descent update rule?

Select all that apply:

- [x] a. $g''(w) = \frac{s}{w^2} + \frac{n-s}{(1-w)^2}$ and update: $w \leftarrow w - \frac{-\frac{s}{w}+\frac{n-s}{1-w}}{\frac{s}{w^2}+\frac{n-s}{(1-w)^2}}$ cross out
- [ ] b. $g''(w) = \frac{s}{w^2} + \frac{n-s}{(1-w)^2}$ and update: $w \leftarrow w + \frac{-\frac{s}{w}+\frac{n-s}{1-w}}{\frac{s}{w^2}+\frac{n-s}{(1-w)^2}}$ cross out
- [ ] c. $g''(w) = -\frac{s}{w^2} + \frac{n-s}{(1-w)^2}$ and update: $w \leftarrow w - \frac{-\frac{s}{w}+\frac{n-s}{1-w}}{-\frac{s}{w^2}+\frac{n-s}{(1-w)^2}}$ cross out
- [ ] d. $g''(w) = \frac{s}{w^2} - \frac{n-s}{(1-w)^2}$ and update: $w \leftarrow w - \frac{-\frac{s}{w}+\frac{n-s}{1-w}}{\frac{s}{w^2}-\frac{n-s}{(1-w)^2}}$ cross out

Question 5

Not yet answered

Marked out of 1.00

Let everything be defined as in the previous question.

What is the closed form solution for

$$w^* = \arg\min_{w \in \mathbb{R}} g(w)$$

Select all that apply:

- [ ] a. $w^* = s/(s-n)$ cross out
- [x] b. $w^* = s/n$ cross out
- [ ] c. $w^* = n/s$ cross out
- [ ] d. $w^* = s/(n-s)$ cross out

Question 6

Not yet answered

Marked out of 1.00

Let $g(w) = w^4 + e^{-w}$ where $w \in \mathbb{R}$.

What is the derivative $g'(w)$ and the first order gradient descent update rule with a constant step size η?

Select all that apply:

- [ ] a. $g'(w) = 4w^3 - e^{-w}$ and update: $w \leftarrow w + \eta(4w^3 - e^{-w})$ cross out
- [ ] b. $g'(w) = 4w^3 + e^{-w}$ and update: $w \leftarrow w - \eta(4w^3 + e^{-w})$ cross out
- [x] c. $g'(w) = 4w^3 - e^{-w}$ and update: $w \leftarrow w - \eta(4w^3 - e^{-w})$ cross out
- [ ] d. $g'(w) = 4w^3 + e^{-w}$ and update: $w \leftarrow w + \eta(4w^3 + e^{-w})$ cross out

Question 7

Not yet answered

Marked out of 1.00

Let everything be defined as in the previous question.

What is the second derivative $g''(w)$ and the second order gradient descent update rule?

Select all that apply:

- [ ] a. $g''(w) = 12w^2 - e^{-w}$ and update: $w \leftarrow w + \frac{4w^3 - e^{-w}}{12w^2 - e^{-w}}$ cross out
- [x] b. $g''(w) = 12w^2 + e^{-w}$ and update: $w \leftarrow w - \frac{4w^3 - e^{-w}}{12w^2 + e^{-w}}$ cross out
- [ ] c. $g''(w) = 12w^2 + e^{-w}$ and update: $w \leftarrow w + \frac{4w^3 - e^{-w}}{12w^2 + e^{-w}}$ cross out
- [ ] d. $g''(w) = 12w^2 - e^{-w}$ and update: $w \leftarrow w - \frac{4w^3 - e^{-w}}{12w^2 - e^{-w}}$ cross out

Question 8

Not yet answered

Marked out of 1.00

Let everything be defined as in the previous question.

For the second order update rule, calculate $w^{(1)}$ if $w^{(0)} = 0$.

Answer: 1

Question **9**

Not yet answered

Marked out of 1.00

Let everything be defined as in the previous question. Change the step size to be calculated using the normalized gradient. For the first order update rule, calculate $w^{(1)}$ if $w^{(0)} = 0$, $\eta = 1$. Only for this problem, set $\epsilon = 0$.

Answer: 1

Question **10**

Not yet answered

Marked out of 1.00

Let $g(\mathbf{w}) = g(w_1, w_2) = w_1^2 w_2^2 + e^{-w_1} + e^{-w_2}$ where $\mathbf{w} \in \mathbb{R}^2$. What is the gradient of $g(w)$ and the first order gradient descent update rule with a constant step size η?

- ○ a. $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \eta\left(2w_1^{(t)}(w_2^{(t)})^2, 2w_2^{(t)}(w_1^{(t)})^2\right)^\top$ cross out
- ◉ b. $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \eta\left(2w_1^{(t)}(w_2^{(t)})^2 - e^{-w_1^{(t)}}, 2w_2^{(t)}(w_1^{(t)})^2 - e^{-w_2^x(t)}\right)^\top$ cross out
- ○ c. $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \eta\left(-2w_1^{(t)}(w_2^{(t)})^2 + e^{-w_1^{(t)}}, -2w_2^{(t)}(w_1^{(t)})^2 + e^{-w_2^{(t)}}\right)^\top$ cross out
- ○ d. $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \eta\left(2w_1^{(t)}(w_2^{(t)})^2 + e^{-w_1^{(t)}}, 2w_2^{(t)}(w_1^{(t)})^2 + e^{-w_2^{(t)}}\right)^\top$ cross out

Clear my choice

Question **11**

Not yet answered

Marked out of 1.00

If $\mathcal{F} \subset \mathcal{G}$, then is it true that $\min_{f \in \mathcal{F}} \hat{L}(f) \geq \min_{g \in \mathcal{G}} \hat{L}(g)$?

◉ True

○ False

Question **12**

Not yet answered

Marked out of 1.00

Consider the setting of polynomial regression. Let $d = 2$, such that $\mathbf{x} = (x_0 = 1, x_1, x_2)$, and $p = 4$, then $\bar{p} = 10$. True or False?

○ True

◉ False

Question 13

Not yet answered

Marked out of 1.00

Let everything be defined as in the previous question. The expression for $\phi_p(\mathbf{x})$ is given by
$\phi(\mathbf{x}) = \left(x_1, x_2, x_1^2, x_1x_2, x_2^2, x_1^3, x_1^2x_2, x_1x_2^2, x_2^3, x_1^4, x_1^3x_2, x_1^2x_2^2, x_1x_2^3, x_2^4\right)$. True or False?

- [x] True
- [ ] False

Question 14

Not yet answered

Marked out of 1.00

Suppose that $\bar{\mathcal{F}}_p = \{f | f : \mathbb{R}^{d+1} \to \mathbb{R}, \text{ and } f(\mathbf{x}) = \log(\phi_p(\mathbf{x})^\top \mathbf{w}), \text{ for some } \mathbf{w} \in \mathbb{R}^{\bar{p}}\}$..
Is it true that $\bar{\mathcal{F}}_1 \subset \bar{\mathcal{F}}_2$?

- [x] True
- [ ] False

Question 15

Not yet answered

Marked out of 1.00

You are predicting house prices. Supose you want to make the irriducible error smaller. If you gather a new feature about houses (that you didn't already have) such as the number of swimming pools in the backyard, is it likely to decrease the irriducible error? True or False?

- [ ] True
- [x] False

Question 16

Not yet answered

Marked out of 1.00

Consider the same setting as the previous problem. The estimation error can be reduced by reducing the number of data points. True or False?

- [ ] True
- [x] False

Question 17

Not yet answered

Marked out of 1.00

Consider the same setting as the previous problem. The approximation error can be reduced by using a larger function class. True or False?

- [x] True
- [ ] False

Question 18

Not yet answered

Marked out of 1.00

You notice your predictor is overfitting. To reduce overfitting, we should make the degree p of the polynomial function class larger. True or False?

- [ ] True
- [x] False

Question 19

Not yet answered

Marked out of 1.00

Suppose that you have a small dataset, but a large function class.
Would the variance be large or small?
Would you expect the bias to be large or small?
Would you expect the predictor $\hat{f}_D$ to be underfitting or overfitting the data or neither?

Select all that apply:

- [ ] a. variance small, bias small, underfit. cross out
- [ ] b. variance small, bias large, overfit. cross out
- [x] c. variance large, bias small, overfit. cross out
- [ ] d. variance large, bias large, overfit. cross out

Question 20

Not yet answered

Marked out of 1.00

Suppose that you have a large dataset, but a small function class, and f_{Bayes} is much more complex than any function in the function class.
Would the variance be large or small?
Would you expect the bias to be large or small?
Would expect the predictor $\hat{f}_D$ to be underfitting or overfitting the data or neither?

Select all that apply:

- [ ] a. variance small, bias small, neither overfitting nor underfitting. cross out
- [x] b. variance small, bias large, underfit. cross out
- [ ] c. variance large, bias large, underfit. cross out
- [ ] d. variance large, bias large, neither overfitting nor underfitting. cross out

Question 21

Not yet answered

Marked out of 1.00

Suppose that you have a large dataset, a small function class $\mathcal{F}$, and $f_{\text{Bayes}} \in \mathcal{F}$.
Would the variance be large or small?
Would you expect the bias to be large or small?
Would expect the predictor $\hat{f}_D$ to be underfitting or overfitting the data or neither?

Select all that apply:

- [ ] a. variance large, bias large, overfitting. cross out
- [ ] b. variance small, bias small, overfitting. cross out
- [x] c. variance small, bias small, neither overfitting nor underfitting. cross out
- [ ] d. variance large, bias large, neither overfitting nor underfitting. cross out

Question 22

Not yet answered

Marked out of 1.00

You are using regularization. You notice you are underfitting.
You should decrease the value of lambda to reduce underfitting and get a smaller test loss. True or False?

- [x] True
- [ ] False

Question 23

Not yet answered

Marked out of 1.00

Suppose you have a dataset $\mathcal{D} = (z_1, \ldots, z_n)$ containing n i.i.d. flips of a coin.
Since the flips are i.i.d. you know they all follow the distribution Bernoulli with parameter (α^*).
However, you do not know what α^* is so you would like to estimate it using MLE.
Which of the following is the maximum likelihood estimate α_{MLE}?

Select all that apply:

- [ ] a. $\alpha_{\mathrm{MLE}} = \frac{1}{n} \sum_{i=1}^{n} \alpha_i$ cross out
- [x] b. $\alpha_{\mathrm{MLE}} = \frac{1}{n} \sum_{i=1}^{n} z_i$ cross out
- [ ] c. $\alpha_{\mathrm{MLE}} = \frac{1}{n} \sum_{i=1}^{n-1} z_i$ cross out
- [ ] d. $\alpha_{\mathrm{MLE}} = \frac{1}{n-1} \sum_{i=1}^{n} z_i$ cross out

Question 24

Not yet answered

Marked out of 1.00

Assume that $Y|X$ follows a Gaussian distribution with mean $\mu = xw_1$ and variance $\sigma^2 = \exp(xw_2)$ for all $x \in \mathbb{R}$ and $\mathbf{w} = (w_1, w_2)$ where $w_1, w_2 \in \mathbb{R}$.
The negative log-likelihood, can be written as follows for a dataset $\mathcal{D} = ((x_1, y_1), \cdots, (x_n, y_n))$:

$$g(\mathbf{w}) = \sum_{i=1}^{n} g_i(\mathbf{w}) \quad \text{where } g_i(\mathbf{w}) = -\ln p(y_i|x_i, \mathbf{w}),$$

where $p(\cdot|\cdot)$ is the density of the above Gaussian distribution. What is partial derivative of g with respect to w_1?

Select all that apply:

- [ ] a. $\frac{\partial g}{\partial w_1} = \sum_{i=1}^{n} \frac{(y_i - x_i w_1)^2}{2\exp(x_i w_2)}$ cross out
- [ ] b. $\frac{\partial g}{\partial w_1} = \sum_{i=1}^{n} \frac{x_i(y_i - x_i w_1)}{\exp(x_i w_2)}$ cross out
- [ ] c. $\frac{\partial g}{\partial w_1} = -\sum_{i=1}^{n} \frac{(y_i - x_i w_1)^2}{\exp(x_i w_2)}$ cross out
- [x] d. $\frac{\partial g}{\partial w_1} = -\sum_{i=1}^{n} \frac{x_i(y_i - x_i w_1)}{\exp(x_i w_2)}$ cross out

Question **25**

Not yet answered

Marked out of 1.00

Let everything be defined as in the previous question.
What is partial derivative of g with respect to w_2?

Select all that apply:

- [ ] a. $\sum_{i=1}^{n}\left(\frac{(y_i-x_iw_1)^2}{2\exp(x_iw_2)}+\frac{x_i}{2}\right)$ cross out
- [ ] b. $\sum_{i=1}^{n}\left(\frac{x_i(y_i-x_iw_1)^2}{2\exp(x_iw_2)}-\frac{x_i}{2}\right)$ cross out
- [ ] c. $\sum_{i=1}^{n}\left(-\frac{(y_i-x_iw_1)^2}{2\exp(x_iw_2)}+x_i\right)$ cross out
- [x] d. $\sum_{i=1}^{n}\left(-\frac{x_i(y_i-x_iw_1)^2}{2\exp(x_iw_2)}+\frac{x_i}{2}\right)$ cross out

Question **26**

Not yet answered

Marked out of 1.00

Let everything be defined as in the previous question.
You want to solve for $\mathbf{w}_{\text{MLE}}$ using gradient descent.
Using the partial derivatives you calculated in the previous quesitons,
what would the gradient update rule look like with a constant step size η?

Select all that apply:

- [ ] a. $w_1 \leftarrow w_1-\eta\sum_{i=1}^{n}\left(\frac{(y_i-x_iw_1)}{\exp(x_iw_2)}\right),\quad w_2 \leftarrow w_2+\eta\sum_{i=1}^{n}\left(\frac{(y_i-x_iw_1)^2}{2}-\frac{x_i}{2}\right)$ cross out
- [ ] b. $w_1 \leftarrow w_1+\eta\sum_{i=1}^{n}\left(\frac{x_i(y_i-x_iw_1)}{\exp(x_iw_2)}\right),\quad w_2 \leftarrow w_2+\eta\sum_{i=1}^{n}\left(\frac{x_i(y_i-x_iw_1)^2}{2\exp(x_iw_2)}-\frac{x_i}{2}\right)$ cross out
- [ ] c. $w_1 \leftarrow w_1-\eta\sum_{i=1}^{n}\left(\frac{(y_i-x_iw_1)}{2}\right),\quad w_2 \leftarrow w_2-\eta\sum_{i=1}^{n}\left(\frac{(y_i-x_iw_1)^2}{2\exp(x_iw_2)}-\frac{x_i}{2}\right)$ cross out
- [x] d. $w_1 \leftarrow w_1-\eta\sum_{i=1}^{n}\left(\frac{x_i(y_i-x_iw_1)}{\exp(x_iw_2)}\right),\quad w_2 \leftarrow w_2-\eta\sum_{i=1}^{n}\left(\frac{x_i(y_i-x_iw_1)^2}{2\exp(x_iw_2)}-\frac{x_i}{2}\right)$ cross out

Last saved at 22:47:28