Week 1 Quiz

Bookmark this page

Check all that apply

1.0/1.0 point (graded)

Check all instances of a supervised learning problem.

- [x] separating spam from non-spam email using the text content of the email
- [ ] organizing people into groups based on a combination of their height, weight and age
- [ ] learning the topics from a corpus of documents
- [x] predicting the presence/absence of a disease based on a blood test

✔

Submit You have used 1 of 2 attempts

Multiple Choice

1/1 point (graded)

The variance of a univariate random variable $x \sim p(x)$ having expectation $\mathbb{E}_q[x] = \mu$ can be written as $\sigma^2 = \mathbb{E}_q[(x-\mu)^2]$. An equivalent equation for calculating this variance is

- ( ) $\sigma^2 = \mathbb{E}_q[x^2] + \mu^2$
- (x) $\sigma^2 = \mathbb{E}_q[x^2] - \mu^2$ ✔
- ( ) $\sigma^2 = \mathbb{E}_q[x^2] + 2\mu^2$
- ( ) $\sigma^2 = \mathbb{E}_q[x^2] - 2\mu^2$

Submit You have used 1 of 1 attempt

True/False

1/1 point (graded)

If $x_1, \ldots, x_n$ are generated independent and identically distributed (i.i.d.) according to the distribution $p(x|\theta)$, then the joint likelihood can be written as

$$p(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} p(x_i|\theta).$$

- False
- True ✔

Submit You have used 1 of 1 attempt

Math Expression Input

1/1 point (graded)

You have data $x_1, \ldots, x_n$ with each $x_i \in \{0, 1\}$. You model this as $x_i \overset{iid}{\sim} Bernoulli(\pi)$. The corresponding joint likelihood is therefore

$$p(x_1, \ldots, x_n|\pi) = \pi^S (1-\pi)^{n-S},$$

where we define $S = \sum_{i=1}^{n} x_i$. Write the maximum likelihood estimate of π.

S/n

✔ **Answer:** s/n

$\frac{S}{n}$

Submit You have used 1 of 2 attempts

Multiple Choice

0/1 point (graded)

You have data pairs $(y_i, x_i)_{i=1:n}$ where $x \in \mathbb{R}^d$ and you perform least squares linear regression to learn a function of the form $y = w_0 + x^T w$. If $w_1 \gg 0$, what does this information immediately tell you about x?

- $x(1)$ is more important than $x(2), \ldots, x(d)$ ✖
- $x(1)$ is directly proportional to y
- $x(1)$ is indirectly proportional to y
- $x(1)$ should be suppressed somehow

Submit You have used 1 of 1 attempt

✖ Incorrect (0/1 point)

Numerical Input

1/1 point (graded)

You have data pairs $(y_i, x_i)_{i=1:n}$ where $x \in \mathbb{R}^{14}$ and you perform least squares linear regression to learn a function of the form $y = w_0 + x^T w$. What is the minimum number of samples required for this to be possible?

15

✔ **Answer:** 15

15

Submit You have used 1 of 1 attempt

Numerical Input

1/1 point (graded)

You want to approximate your data as a quadratic function using least squares with a polynomial of the form $y = w_0 + \sum_{r=1}^{p} w_r x^p$. What value should you set p to?

2

✔ **Answer:** 2

2

Submit You have used 1 of 1 attempt

✔ Correct (1/1 point)

Numerical Input

1/1 point (graded)

You have data pairs $(y_i, x_i)_{i=1:17}$ where $x \in \mathbb{R}$ and you perform least squares polynomail regression to learn a function of the form $y = w_0 + \sum_{r=1}^{p} w_r x^p$. What is the maximum value you can set p to?

16

✔ **Answer:** 16

16

Submit You have used 1 of 1 attempt

✔ Correct (1/1 point)

Dropdown

1/2 points (graded)

You have n pairs of observations (y_i, x_i) where $x \in \mathbb{R}^{d+1}$ and the first dimension of x equals 1. You perform least squares on $y \approx x^T w$ to learn w. From the lectures we discussed how, using the coefficients w_{LS}, you can think of the errors in two ways.

1. When thinking of $y_i \approx x_i^T w_{LS}$ for $i = 1, \ldots, n$, the errors $y_i - x_i^T w_{LS}$ are ____ to the ____-dimensional hyperplane in $\mathbb{R}^d$.

not perpendicular, d ▼ ✖ **Answer:** not perpendicular, $d - 1$

2. When we think of $y \in \mathbb{R}^n$ and $\hat{y} = Xw_{LS}$, then $y - Xw_{LS}$ creates an error vector orthogonal to ____.

y^ ▼ ✔ **Answer:** $\hat{y}$

Submit You have used 1 of 1 attempt

✱ Partially correct (1/2 points)