

Question 2 – MAP

Suppose the prior for some parameter β is $N(5, \sigma_p^2)$, and given data the likelihood is proportional to $N(10, \sigma_l^2)$. What are possible values for the maximum a posteriori estimation of β if we only know that $\sigma_p, \sigma_l > 0$?

- ☐ $\hat{\beta} = 5$
- ☒ $\hat{\beta} = 6.5$
- ☐ $\hat{\beta} = 11$
- ☐ $\hat{\beta} = 10$

11.06.2024

3

Question 12 – Linear Regression

You are doing linear regression with multiple predictors and an intercept by minimizing the OLS loss. Which of the following statements is true?

- ☒ If a feature x_i other than the intercept feature is constant, there is no unique optimal estimate $\hat{\beta}_i$.
- ☒ If one feature is twice another $x_i = 2x_j$, there is no unique solution for the optimal parameters $\hat{\beta}_i$ and $\hat{\beta}_j$.
- ☐ If one feature's values are always larger than another feature's values $[x_i]_j > [x_j]_j$ for all $j = 1, \dots, n$, then $\hat{\beta}_i > \hat{\beta}_j$.
- ☒ Adding a new feature x_{new} that is uncorrelated to all previous features does not change the optimal parameter estimates of old features.

11.06.2024

4

Question 9 – Fitting Linear Regression

Which of the following models can be fit with linear regression?

- ☒ $y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2^2$
- ☒ $y = \beta_0 + \beta_1^3 x_1$
- ☐ $y = \beta_0 + \sin(\beta_1 x_1)$
- ☒ $y = \beta x^4$

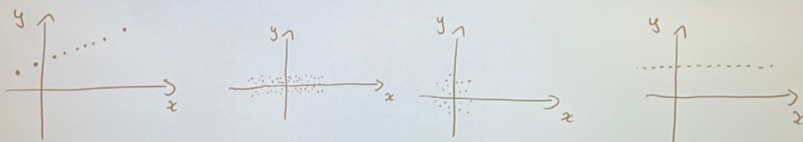
Question 1 – Cross Validation

You do $K = 5$ fold cross validation to select the best of 7 possible values for a hyperparameter. After cross-validation you want to use the best hyperparameter value for prediction. Your dataset has $n = 100$ elements. Which is correct:

- ☒ You train a total of 36 models and each training set has size 80.
- ☐ You train a total of 6 models and your training sets have size 20.
- ☐ You train a total of 16 models and your training sets have size 80.
- ☐ This is not possible, K needs to be larger than the number of possible hyperparameter values.

Question 13 – MLE

You compute the maximum likelihood estimation $\hat{\theta}$ in probabilistic linear regression. For which of the following training datasets will $\hat{\theta} \approx 0$?



Question 10 – Gradient Descent

Which of the following statements about gradient descent are correct?

- ☐ Gradient descent cannot be applied if the loss function has more than one minimum.
- ☒ If the step size is too large, the loss may increase.
- ☐ Gradient descent always converges to the minimum that is closest to the starting value.
- ☒ While not at a minimum, there is always a step size small enough so that the loss decreases.

Question 4 – Cross Validation

We are trying to validate the model using 5-fold cross-validation, but the error is suspiciously small. *In which line is the error?*

```
# X contains two input features, y contains the target values
X = np.column_stack([feat1, feat2]) # shape (n_samples, 2)
y = price # shape (n_samples,)
cv = KFold(n_splits=5, shuffle=True, random_state=0)
# KFold splits the dataset into 5 train/test partitions
for train_idx, test_idx in cv.split(X):
    model = LinearRegression().fit(X[train_idx], y[train_idx])
    print(f"Test error: {mean_squared_error(y[train_idx], model.predict(X[train_idx]))}")
```

should use test_idx

11.06.2024

9

Question 14 – Model Fitting

Which of the following statements are true?

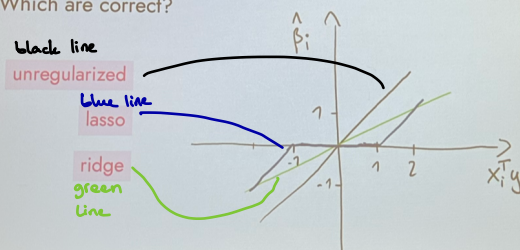
- ☒ Model parameters can be fit to training data by minimizing a loss.
- ☒ Model hyperparameters should be selected on a validation set.
- ☐ Hyperparameters have larger values than parameters.
- ☒ Without a test set you cannot estimate the performance of your model on unseen data.

11.06.2024

10

Question 16 – Regularization

Suppose that all features are orthonormal to each other, i.e., $x_i^T x_j = 0$ if $i \neq j$ and $\|x_i\| = 1$ for all i . The following plot shows the estimation for the i -th entry of β as a function of $x_i^T y$ for an unregularized, ridge-regularized, and lasso-regularized linear regression model. Which are correct?

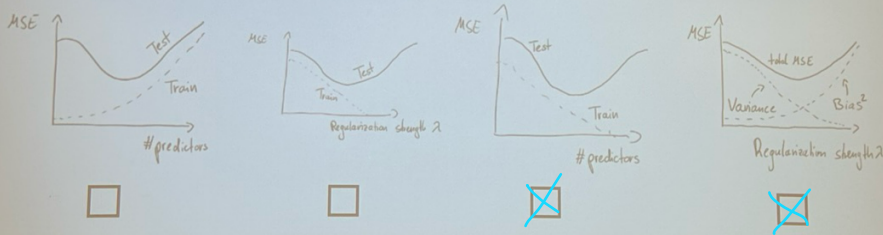


11.06.2024

11

Question 3 – Bias-Variance Trade-Off

The following four sketches show MSE behavior in the context of a bias variance trade-off. Select the ones that show correct behavior.



Question 5 – Linear Regression

Which lines in this code can be omitted without changing the mse result?

```
import numpy as np
n = 10
X = np.random.randn(n, 3)
beta = np.array([0.3, 5, 2])
y = X @ beta + np.random.randn(n)*0.01
XtX = X.T @ X
beta_hat = np.linalg.inv(X.T @ X) @ X.T @ y
y_pred = X @ beta_hat
residuals = y - y_pred
mse = np.mean((y - X @ beta_hat)**2)
```

11.06.2024

15

Question 8 – Derivation of Variance

The variance of a random variable X is $\text{Var}(X) = \mathbb{E}[(X - \mu)^2]$, where $\mu = \mathbb{E}[X]$. Which of the following is an equivalent expression for the variance?

- ☒ $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- ☐ $\text{Var}(X) = \mathbb{E}[X]^2 - \mathbb{E}[X^2]$
- ☐ $\text{Var}(X) = \mathbb{E}[X - \mathbb{E}[X]]$
- ☐ $\text{Var}(X) = \mathbb{E}[X]^2$

11.06.2024

16

Question 7 – MSE

How do we fix the computation of the mean squared error (MSE) in the code below?

```
import numpy as np
y = np.array([1.0, 2.0, 3.0])
y_pred = np.array([0.9, 2.1, 2.9])
mse = np.sqrt(np.mean(y_pred - y))
```

- ☐ Should remove the square root.
- ☒ Should use $\text{np.mean}((y_{\text{pred}} - y)**2)$.
- ☐ Should divide by $\text{len}(y) - 1$.
- ☐ Should replace np.sqrt with np.std .

11.06.2024

Question 6 – Linear Regression

The code produces `NotFittedError: This LinearRegression instance is not fitted yet...` What should you do?

```
from sklearn.linear_model import LinearRegression
import numpy as np
X = np.random.randn(30, 2)
y = np.random.randn(30)
model = LinearRegression()
y_pred = model.predict(X)
```

- ☒ Must call `model.fit(X, y)` before predict.
- ☐ Intercept must be disabled.
- ☐ Should import Ridge instead.
- ☐ Model requires reshaping y.

Question 15 – Optimization

Match these concepts in optimization with their corresponding mathematical expression.

OLS

$$\min_{\beta} \|y - X\beta\|_2^2$$

Ridge

$$\min_{\beta} (\|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2)$$

Lasso

$$\min_{\beta} (\|y - X\beta\|_2^2 + \lambda \|\beta\|_1)$$

Gradient ..

$$\beta^{(t+1)} = \beta^{(t)} - \eta \nabla_{\beta} \|y - X\beta^{(t)}\|_2^2$$

Ridge Regression

Lasso

OLS

Gradient Descent

Question 11 – Gradient Update

Which of these is the correct gradient descent update for the loss $L(\beta) = 3\beta^2 + 3\beta + 4$?

- ☒ $\beta \leftarrow \beta - \eta (6\beta + 3)$
- ☐ $\beta \leftarrow \beta + \eta (6\beta + 3)$
- ☐ $\beta \leftarrow \beta - \eta (3\beta^2 + 3\beta + 4)$
- ☐ $\beta \leftarrow \beta + \eta 6\beta$