DATA 100: Vitamin 11 Solutions

November 1, 2019

1 Mean Squared Error

Suppose we have data $(x_i, y_i)_{i=1}^n$, where $\vec{x}$ is the independent variable, and $\vec{y}$ is the dependent variable. We fit a model to the data, and obtain predicted values of the dependent variable, denoted $\vec{\hat{y}}$. We've seen the mean squared error (MSE) in many forms. Which of the following equations correspond to the MSE of the dependent variable and its predicted values?

- ☑ $\frac{1}{n}\|\vec{\hat{y}} - \vec{y}\|_2^2$
- ☑ $\frac{1}{n}\|\vec{y} - \vec{\hat{y}}\|_2^2$
- ☐ $\frac{1}{n}\|\vec{y} - \vec{\hat{y}}\|_2$
- ☑ $\frac{1}{n}\|\vec{y} - \vec{\hat{y}}\|^2$
- ☐ $(y - \hat{y})^2$
- ☑ $\frac{1}{n}\sum_{i=1}^n (y_i - \hat{y}_i)^2$

Explanation: These answers follow directly from the definition of norms: $\|\vec{z}\|_p = (\sum_{i=1}^n |z_i|)^{\frac{1}{p}}$, $\vec{z} = (z_1, \ldots, z_n)^\top$.

2 Penalized Linear Regressions

Fill in the blanks:
We fit a penalized linear regression model without an intercept to some data by minimizing the MSE. Since our regularization term is the ____, we are fitting a LASSO regression. For extremely large penalty hyperparameters, we expect the coefficients of our model to be ____.

- ☐ sum of squares, very close to 0
- ☐ sum of absolute values, very close to 0
- ☐ sum of squares, exactly 0

☑ sum of absolute values, exactly 0

Explanation: Since we are fitting a LASSO regression, the penalization term is the sum of absolute values. As the hyperparameter dictating the regularization increases to infinity, the coefficients of our model go to zero.

3 Global Minimums

For which of the following functions is gradient descent guaranteed to find a global minimum? (Hint: plot these functions)

☑ $f(x) = x^2 + 100$

☐ $f(x) = x^3 - 10x^2 - 4x + 10$

☐ $f(x) = -2x^5 + 8x^4$

☑ $f(x) = x^6 - \sqrt{x}$

☑ $f(x) = log(x) + x^4 + \frac{1}{x^2}$

Explanation: Gradient descent will find the global minimum of convex functions. Plot these functions to determine which ones are convex.

4 Gradient Descent

Which of the following statements about gradient descent (GD) are true?

☑ GD is a numerical method

☐ GD provides analytic solutions to risk minimization problems

☐ GD permits the risk minimization of any combination of risk function and model.

☑ If the gradient of the risk of a model exists, then we can minimize the risk using GD.

Explanation: GD allows us to minimize the risk of a model numerically, but only if the gradient of the model's risk exists.

5 Logistic Regression

Fill in the blanks:
Logistic regression models can be used for ____, meaning they are used to predict ____ variables.

- ☐ classification, quantitative
- ☑ classification, categorical
- ☐ regression, categorical
- ☐ clustering, discrete

Explanation: Logistic regression is a model that is used in classification settings, meaning they are used to predict categorical variables.