DATA 100: Vitamin 8 Solutions

March 23, 2019

1 Regression Models

1.1 Definition and Motivation

Fill in the blanks: Suppose we draw a sample of random variable Y and a set of potentially related features X, and we are interested in quantifying the relationship between Y and X. A natural way to accomplish this is through a regression function, which gives the ____ of Y given X. Generally, ____

- ☐ conditional value, there are infinitely many such functions
- ☑ conditional expected value, there are infinitely many such functions
- ☐ conditional value, such a function is unique
- ☐ conditional expected value, such a function is unique

Explanation: A natural way of investigating the relationship between some variable Y and a set of covariates X is through a regression function, which is defined as the conditional expected value of Y given X. In general, there are an infinite number of such regression functions.

1.2 Linear Regression Functions

Why are linear regression functions used so often in practice?

- ☑ They are computationally easy to find.
- ☑ They are interpretable.
- ☐ They make assumptions that hold generally in practice.
- ☑ They can accommodate both quantitative and qualitative covariates.

Explanation: Option three is false, since the key assumption that is required to use linear regression models for inference is often violated. n the case of the ordinary least squares estimator, this assumption is that the residuals are independently and idendically normally distributed across all observations. The property that $Y_n|X_n \sim N(X_n\beta, \sigma I_n)$ is a direct result of this assumption on the residuals.

1.3 Identify the Linear Regression Models

Let $E[Y|X] = \theta(X)$ and let β be a vector of coefficients. Which of the following $\theta(X)$ are linear regression models?

- [x] β_1
- [x] $\beta_1 + \beta_2 X_1 X_2 + \beta_3 X_3^{X_4}$
- [ ] $\exp(\beta_1 X_1 + \beta_2 X_2)$
- [x] $\beta_1 \ln(X_1) + \beta_2 \exp(X_2)$
- [ ] None of these

Explanation: Options 1,2 and 4 are linear regression models since each of their terms are linear with respect to their β coefficients.

2 Gradient Descent Algorithms

2.1 Properties and Uses

Gradient descent algorithms:

- [x] Can be used to solve optimization problems that have no analytical solutions.
- [ ] Always find the global optimum if it is unique.
- [x] Can require tuning algorithm parameters.

Explanation: Gradient descent algorithms require a judicious choice of tuning parameters, such as the learning rate, as well as starting values and stopping rules. They are used to solve optimization problems when analytical solutions are impossible or intractable.

2.2 Convexity

Which of the following reasons are convex loss functions favored over non-convex loss functions?

- ☑ A local optimum of a convex function is always a global optimum.
- ☐ The risk minimizer for a convex loss always has an analytical solutions.
- ☐ Gradient descent with a convex loss does not require a stopping criterion.

Explanation: Option two is incorrect since the convexity of a function does not guarantee that the global minima can be found analytically. See the Huber loss function. Option three is incorrect since iterative algorithms require a stopping criterion.