Homework 5

ECE4530J - Decision Making in Smart Cities Summer 2022

* Name: Huang Yucheng ID: 519021910885

Problem 1

Consider a linear regression model with the hypothetical relation

$$y = \beta^T x.$$

a) Given one practical example which can be well modeled by such a linear model. Clearly define the predictors and the response. Explain why.

b) Given one practical example which cannot be well modeled by such a linear model. Clearly define the predictors and the response. Explain why not.

c) Given one practical example which can be approximately modeled by such a linear model, with possibly significant error sometimes. Clearly define the predictors and the response. Explain why.

Answer:

a) EV charging demand.

response y: demand for EV charging

predictor x: land use, population density, neighborhood

Explain: When is linear relation a good assumption? - Response variable is monotonic in independent variable, The experiments use $x_1, x_2, \ldots, x_n$ as the values of the dependent variables. The experiments lead to the following results: $y_1, y_2, \ldots, y_n$. Hypothesize that $y = \beta_0 + \beta_1 x + e$, and we can get the model.

b) Masking Problem, Linear discriminant analysis

response y: Classification of multi set of points.

predictor x: A vector, including coordinates and $\hat{\pi}_k = N_k/N$, where N_k is the number of class- k observations

Explain: Logit function

$$\Pr\{G = k \mid X = x\} = \frac{\exp\left(\beta_{k0} + \beta_k^T x\right)}{\sum_{l=1}^{K} \exp\left(\beta_{l0} + \beta_l^T x\right)}$$

Classification

$$G(x) = \arg\max_k \frac{\exp\left(\beta_{k0} + \beta_k^T x\right)}{\sum_{l=1}^{K} \exp\left(\beta_{l0} + \beta_l^T x\right)}$$

Then, it is a logistic regression, not a linear model.

c) house price forecast

response y: House Price

predictor x: Location, garage, size, basement, construction time

Explain: For housing prices, there are many factors that depend on the above, but the location of the house is the most influential factor, because it includes multiple relationships such as environment, temperature, traffic, friends, etc., so it is basically a linear relationship, and other small items can be used as Smaller noise terms are ignored.

Problem 2

Suppose that we use smart meters to infer the usage of home appliances.

a) What data does a smart meter measure?

b) Why we need to retrieve "signatures" from the data rather than directly using the original data for the inference?

c) Suppose that we use a linear function

$$G_k(x) = \beta^T x - \gamma_k$$

to determine whether appliance k is "on" or "off". That is, we classify appliance k to be "on" if and only if $G_k(x) > 0$. Use 1-2 sentences to describe how to obtain the coefficients β via linear regression.

d) Does the linear regression approach in part (c) always work for general classification problems? Why or why not?

Answer:

a) It will measure the energy consumption data of a whole house, and also track major home appliances based on the total signal collected by utility meters

b) Since the signature will represent different appliances. For example, a washer usually follows the following operating modes: water-fill, immerse, rinse, drainage, A TV set may experience a falling spike at moments of switching channels. It is enough to track the appliance by signatures. If we use original data, it will take a lot of time and money.

c) γ_k is a threshold for type k appliance

β^T and γ_k determined by linear regression and optimized by validation set

Suppose that we have n observations

$$\text{RSS} = \sum_{i=1}^{n} \sum_{k=1}^{K} (y_{i,k} - (\beta_{k,0} + \beta_{k,1}x_1 + \beta_{k,2}x_2 + \cdots + \beta_{k,p}x_p))^2 = \left(Y - X^T B\right)^T \left(Y - X^T B\right)$$

$$\text{Minimize RSS } \hat{B} = \left(X^T X\right)^{-1} X^T Y$$

d) LR not always work

- It seems that $\hat{y}_k \approx \Pr\{G = k \mid X = x\}$

- However, $\hat{y}_k = \beta_{k,0} + \beta_{k,1}x_1 + \beta_{k,2}x_2 + \cdots + \beta_{k,p}x_p$ can be < 0 or > 1, which is not allowed for probabilities

- Another serious problem: masking

Problem 3

Answer the following questions on neural networks.

a) What is a deep neural network?

b) Why this class of machine learning algorithms are called "neural networks"?

c) What is an activation function?

d) (bonus) Suppose that you are using a neural network (NN) for an engineering task. How would you determine the structure of the NN?

Answer:

a) The units in the middle of the network, computing the derived features Z_m, are called hidden units because the values Z_m are not directly observed.

   In general there can be more than one hidden layer: deep neural networks.

b) - The model was first derived for human brain
   - Each hidden unit represents a neuron
   - Each connection represents a synapsis
   - Each input represent an external signal (sense)
   - A neuron is activated if the total signal passed to it exceed a certain threshold (sigmoid function).

c) A neuron is activated if the total signal passed to it exceed a certain threshold (sigmoid function).

$$Z_m = \frac{1}{1 + \exp\left(\alpha_{0m} + \alpha_m^T X\right)}$$

   Also called activation function

d) Since NN is a two-stage regression or classification model .Instead of directly feeding predictor variables into a regression/classification function, we put a set of intermediate derived features or hidden units in between.

   The NN is essentially a nonlinear regression. We have a p-dimensional input vector $X = \begin{bmatrix} X_1 & X_2 & \ldots & X_p \end{bmatrix}^T$

   Our objective is to predict a K-dimensional output vector $Y = \begin{bmatrix} Y_1 & Y_2 & \ldots & Y_K \end{bmatrix}^T$ from X.

   Hidden units: an M-dimensional vector $Z = \begin{bmatrix} Z_1 & Z_2 & \ldots & Z_M \end{bmatrix}^T$

$$\begin{aligned} Z_m &= \frac{1}{1 + \exp\left(\alpha_{0m} + \alpha_m^T X\right)'} \\ m &= 1, 2, \ldots, M \\ Y_k &= \beta_{0k} + \beta_k^T Z, Z = 1, \ldots, K \end{aligned}$$

   Then the NN structure becomes a regression problem.