DBDA.X408(33) Homework 1

Bill Chen

April 2023

Note

(1) These questions require thought but do not require long answers. Please be as concise as possible. (2) If you have a question about this homework, we encourage you to post it on our Canvas forum or email it to the instructor xchen375@ucsc.edu.

In the first two lectures, you have seen how to fit a linear function of the data for the regression problem. In this question, we will see how linear regression can be used to fit non-linear functions of the data using feature maps. We will also explore some of its limitations, for which future lectures will discuss fixes.

(a) Learning degree-3 polynomials of the input:

Suppose we have a dataset $\{(x^{(i)}, y^{(i)})\}_{i=1}^{n}$ where $x^{(i)}, y^{(i)} \in \mathbb{R}$. We want to find a third-degree polynomial $h_\theta(x) = \theta_3 x^3 + \theta_2 x^2 + \theta_1 x^1 + \theta_0$ to the dataset. The key observation here is that the function $h_\theta(x)$ is still linear in the unknown parameter θ, even though it's not linear in the input x. This allows us to convert the problem into a linear regression problem as follows.

Let $\phi : \mathbb{R} \to \mathbb{R}^4$ be a function that transforms the original input x to a 4-dimensional vector defined as

$$\phi(x) = \begin{bmatrix} 1 \\ x \\ x^2 \\ x^3 \end{bmatrix} \in \mathbb{R}^4 \tag{1}$$

Let $\hat{x} \in \mathbb{R}^4$ be a be a shorthand for $\phi(x)$, and let $\hat{x}^{(i)} \triangleq \phi(x^{(i)})$ be the transformed input in the training dataset. We construct a new dataset $\{(\phi(x^{(i)}), y^{(i)})\}_{i=1}^{n} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{n}$ by replacing the original $x^{(i)}$'s by $\hat{x}^{(i)}$'s. We see that fitting $h_\theta(x) = \theta_3 x_3 + \theta_2 x_2 + \theta_1 x_1 + \theta_0$ to the old dataset is equivalent to fitting a linear function $h_\theta(hatx) = \theta_3 hatx_3 + \theta_2 hatx_2 + \theta_1 hatx_1 + \theta_0$ to the new dataset because

$$h_\theta(x) = \theta_3 x_3 + \theta_2 x_2 + \theta_1 x_1 + \theta_0 = \theta_3 hatx_3 + \theta_2 hatx_2 + \theta_1 hatx_1 + \theta_0 \tag{2}$$

In other words, we can use linear regression on the new dataset to find parameters $\theta_0, ..., \theta_3$.

Please write down

1. the objective function $J(\theta)$ of the linear regression problem on the new dataset
2. the update rule of the batch gradient descent algorithm for linear regression on the dataset

Terminology: In machine learning, ϕ is often called the feature map, which maps the original input x to a new set of variables. To distinguish between these two sets of variables, we will call x the input **attributes**, and call $\phi(x)$ the **features**. (Unfortunately, different authors use different terms to describe these two things. In this course, we will do our best to follow the above convention consistently.)

(b) Coding question: degree-3 polynomial For this question, use the *house.csv* file provided in the week 2 folder on the course website. This csv file contains a small dataset with multiple feature variables (x's), such as areas, and one target variable (y's) `median_house_value`. Pick a feature variable that you prefer to fit a linear model. And please provide the analysis and proof for the rationale why you choose this feature variable.

Using the formulation of the previous sub-question, implement linear regression with normal equations using the feature map of degree-3 polynomials.

Create a scatter plot of the training data, and plot the learned hypothesis as a smooth curve over it. Submit the plot in the write-up as the solution for this problem.

(c) Coding question: degree-k polynomial regression Now we extend the idea above to degree-k polynomials by considering $\phi : \mathbb{R} \to \mathbb{R}^k + 1$ to be

$$\phi(x) = \begin{bmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^k \end{bmatrix} \in \mathbb{R}^{k+1} \tag{3}$$

Follow the same procedure as the previous sub-question, and implement the algorithm with $k = 3, 5, 10, 20$. Create a plot similar to the previous question, and include the hypothesis curves for each value of k with a different color. Include a legend in the plot to indicate which color is for which value of k.

Submit the plot in the write-up as the solution for this problem. Observe how the fitting of the training dataset changes as k increases. Briefly comment on your observations in the plot.

(d) Coding question: solution for overfitting During the lecture, you learned about overfitting. Using a high degree of the polynomial from the previous question will overfit the data. Please provide a solution for this issue but still maintain high model accuracy. Submit the plot in the write-up as the solution for this problem.