

National University of Singapore
School of Computing
CS2109S: Introduction to AI and Machine Learning
Semester I, 2023/2024

Tutorial 4
Decision Trees and Linear Models

These questions will be discussed during the tutorial session. Please be prepared to answer them.

Summary of Key Concepts

In this tutorial, we will discuss and explore the following learning points from Lecture:

1. Decision Trees
 - (a) Decision Tree Learning Algorithm
2. Linear Models
 - (a) Normal Equation
3. Cost functions
4. Gradient Descent
 - (a) Effect of learning rates

Problems

1. Decisions That Matter.

The loans department of DBN (Development Bank of NUS) wanted to use a decision tree to help them make decisions for new applicants. DBN has the following past loan processing records, each containing an applicant's income, credit history, debt, and the final approval decision. Details are shown in Table 1.

Income	Credit History	Debt	Decision
Over 10k	Bad	Low	Reject
Over 10k	Good	High	Approve
0 - 10k	Good	Low	Approve
Over 10k	Good	Low	Approve
Over 10k	Good	Low	Approve
Over 10k	Good	Low	Approve
0 - 10k	Good	Low	Approve
Over 10k	Bad	Low	Reject
Over 10k	Good	High	Approve
0 - 10k	Bad	High	Reject

Table 1: Loan Processing Outcomes

- (a) Construct the best decision tree to classify the final outcome (Decision) from the three features Income, Credit History, and Debt.

- (b) It turns out in the labeling process, one of the data is mislabeled. The data in table 1 is now altered into table 2.

Income	Credit History	Debt	Decision
Over 10k	Bad	Low	Approve
Over 10k	Good	High	Approve
0 - 10k	Good	Low	Approve
Over 10k	Good	Low	Approve
Over 10k	Good	Low	Approve
Over 10k	Good	Low	Approve
0 - 10k	Good	Low	Approve
Over 10k	Bad	Low	Reject
Over 10k	Good	High	Approve
0 - 10k	Bad	High	Reject

Table 2: Loan Processing Outcomes (Noisy)

Construct the best decision tree that classifies the data in table 2. Justify this decision tree and show why it performs the best by calculating the information gain values and remainders at each stage.

(Note: $\log_2 \frac{x}{y} = \log_2 x - \log_2 y$, $\log_2 1 = 0$, $\log_2 2 = 1$, $\log_2 3 = 1.585$, $\log_2 4 = 2$, $\log_2 5 = 2.322$, $\log_2 6 = 2.585$, $\log_2 7 = 2.807$, $\log_2 8 = 3$, $\log_2 9 = 3.170$, $\log_2 10 = 3.322$)

- (c) What is the decision made by the decision tree in part (b) for a person with an **income over 10k**, a **bad credit history**, and **low debt**?
- (d) The situation in part (c) demonstrates inconsistent data in the decision tree i.e. the attribute does not provide information to differentiate the two classes. In practice, it is usually left undecided. What are some ways to mitigate inconsistent data?
- (e) Let's consider a scenario where you desire a Decision Tree with each leaf node representing a minimum of 3 training data points. Derive the tree by pruning the tree you previously obtained in part (b). Which data(s) do you think are likely outlier(s)?

2. Linear Regression Model Fitting.

You are given several data points as follows:

x_1	x_2	x_3	y
6	4	11	20
8	5	15	30
12	9	25	50
2	1	3	7

- (a) Apply the Normal Equation formula to obtain a linear regression model that minimizes MSE of the data points.

$$w = (X^T X)^{-1} X^T Y$$

- (b) Normal Equation needs the calculation of $(X^T X)^{-1}$. But sometimes this matrix is not invertible. When will that happen, and what should we do in that situation?

3. Examining Cost Functions.

For Linear Regression, there are two popular cost functions,

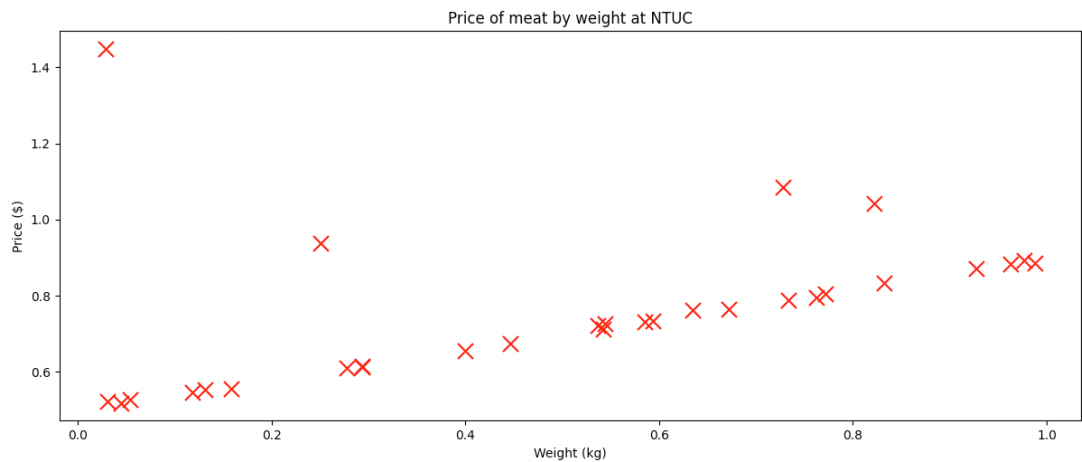
Mean Squared Error:

$$L(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2 \quad (1)$$

and **Mean Absolute Error:**

$$L(y, \hat{y}) = \frac{1}{2}|y - \hat{y}| \quad (2)$$

- (a) Given the scatter plot of a dataset containing the actual weight of meat at NTUC (x) and its price (y), justify your choice of cost function for this problem.

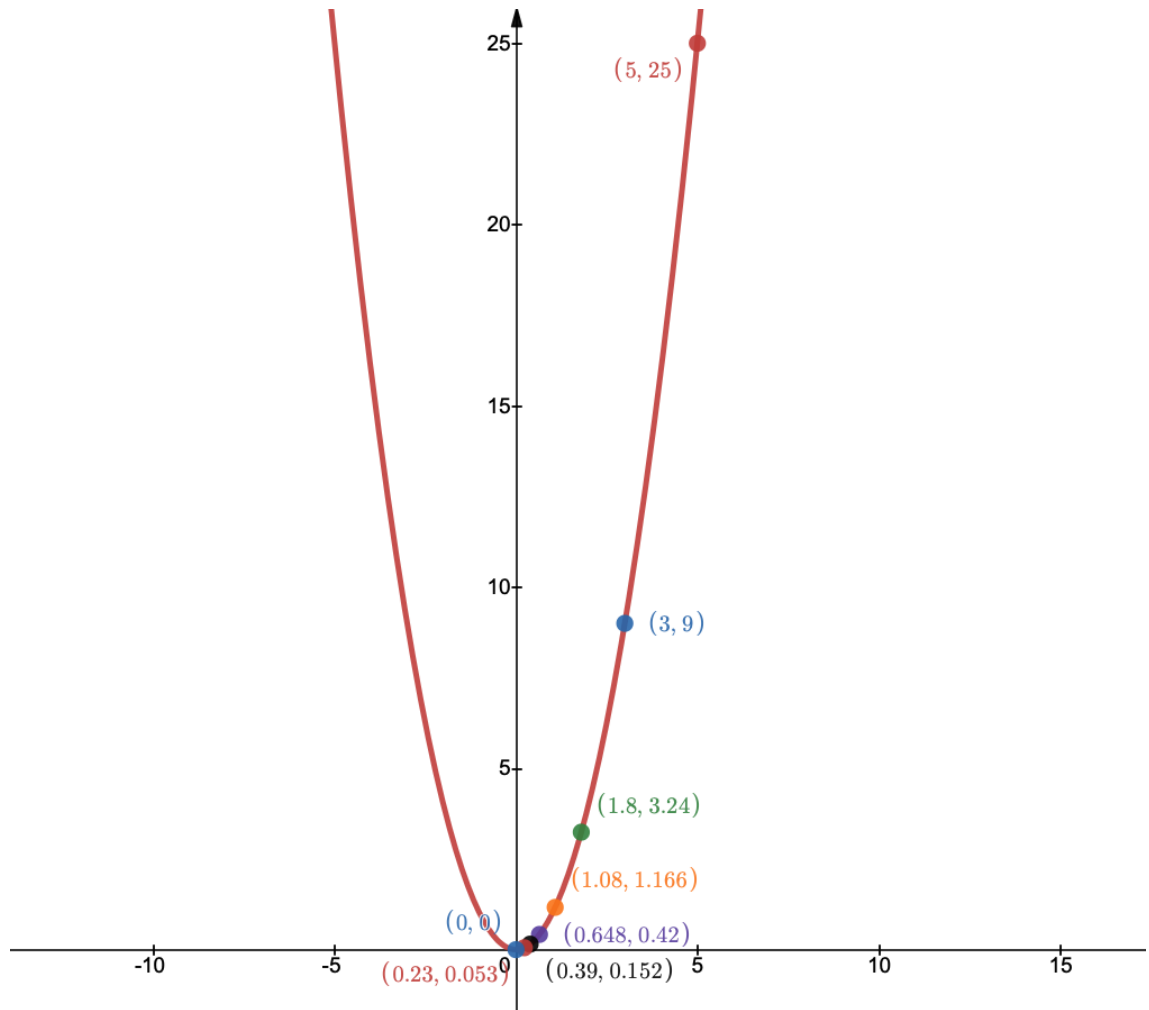


- (b) Can you provide examples of cost functions that are better suited to handle outliers more effectively?

4. Choosing Learning Rates.

- (a) Given a simple function $y = x^2$, we know the gradient is $\frac{dy}{dx} = 2x$. As such, the minimum of this function is 0.

Start from the point $a = (5, 25)$, and using the different α values from $\{10, 1, 0.1, 0.01\}$, perform Gradient Descent on values of x . Record the value of $a = (x, y)$ (where $y = x^2$) over **5 iterations** for each learning rate in tabular format. Observe the oscillations of the value and the convergence to $(0, 0)$. An example is shown below for $\alpha = 0.2$ over 7 steps:



- (b) During the course of training for a large number of epochs/iterations, what can be done to the value of the learning rate α to enable better convergence?

Hint: Consider the equation for gradient descent $w_j = w_j - \alpha \frac{\partial J(w_1, w_2, \dots)}{\partial w_j}$