**Week 1**

* **Introduction:**
  + What is Machine Learning?
    1. (Arthur Samuel, 1959) The science of getting computers learned without being explicitly programmed. (Google search, spam detector)
    2. Tom Mitchell provides a more modern definition: "A computer program is said to learn from **experience E** with respect to some class of **tasks T** and **performance measure P**, if its performance at tasks in T, as measured by P, improves with experience E."
* Grew out of Artificial Intelligence.
* New capability for computers.
  + Examples of machine learnings
    - Database mining

Web click data, medical records, biology, engineering.

* + - Applications can’t be processed by hand

Autonomous helicopter, Natural Language Processing (NLP), handwriting recognition, computer vision.

* Self-customized programs

Amazon, Netflix recommendations

* Understanding human learning
  + Types of machine learning:
    1. **Supervised learning (“right answers” given):**

In supervised learning, we are given a data set and already know what our correct output should look like, having the idea that there is a relationship between the input and the output. Supervised learning problems are categorized into "regression" and "classification" problems.

1. **Regression problems:**

In a regression problem, we are trying to **predict results within a continuous output**, meaning that we are trying to map input variables to some continuous function.

* + - * + House prices **prediction** based on actual house sizes & sale prices.
        + Large inventory of **identical** items, **predict** how many of them will be sold in the next two months.

1. **Classification problems:**

In a classification problem, we are instead trying to **predict results in a discrete output**. In other words, we are trying to map input variables into discrete categories.

* + - * + **Decide** whether a tumor is malignant or benign.
        + **Decide** whether each individual customer account has been hacked/compromised.
    1. **Unsupervised learning:**

Unsupervised learning allows us to approach problems with little or no idea what our results should look like. We can derive structure from data where we don't necessarily know the effect of the variables.

1. **Clustering:**

Take a collection of 1,000,000 different genes, and find a way to automatically group these genes into groups that are somehow similar or related by different variables, such as lifespan, location, roles, and so on.

1. **Non-clustering:**

The "Cocktail Party Algorithm", allows you to find structure in a chaotic environment. (i.e. identifying individual voices and music from a mesh of sounds at a cocktail party).

* + 1. Reinforcement learning:

Concerned with how intelligent agents ought to taka actions in an environment in order to maximize the notion of cumulative reward.

* **Model and Cost Function:**
  + Diagram

    Description automatically generatedModel Representation (supervised learning):
  + Cost function:

We can measure the accuracy of our hypothesis function by using a **cost function.**

A picture containing text, clock, watch, gauge

Description automatically generatedExample:

This function is otherwise called the "Squared error function", or "Mean squared error". The mean is halved as a convenience for the computation of the gradient descent, as the derivative term of the square function will cancel out the term.

**\*The cost function is a function of and .**

* + Intuition of Cost function:

To minimize the average squared vertical distances of the scattered points.

* **Parameter learning:**
  + Chart, surface chart

    Description automatically generatedGradient descent:

\*Different start point may end up with different result.

Finding the local minimum by updating the parameters of the cost function. Repeat the algorithm until convergence.

A picture containing logo

Description automatically generated**: Learning rate**

Text

Description automatically generated with low confidenceCorrect procedure: simultaneously update the parameters and

* + Some characteristics and intuition of gradient descent:
    1. **It is self-adjusting:**

When derivatives are positive (cost is increasing as a parameter increases), the procedure will decrease the parameter.

When derivatives are negative (cost is decreasing as a parameter increases), the procedure will increase the parameter.

* + 1. **Diagram

       Description automatically generated with medium confidenceIf (learning rate) is too large, it may fail to converge, or even diverge.**
    2. **The derivative term approaches 0 as we approach the bottom of our convex function, which is the intuition behind convergence.**
  + **It turns out that there is only one global optimum for cost function of linear regression, and the technical term for the bow shaped function is convex function.**
  + **“Batch” Gradient Descent:**

Each step of gradient descent uses all the training examples.

* **Linear Algebra Review**
  + **Matrix**: Rectangular (2-dimentional) array of numbers.

Definition: number of rows (横行) x number of columns (竖列).

* + **Vector:** An nx1 matrix. (One column)
  + Addition and subtraction: They are element-wise. You can only add matrices with same dimension.
  + Scalar Multiplication (multiply the matrix with a number): just multiply each element in the matrix with the same value.
  + Matrix vector multiplication:

The multiplication can be applied only to (m x n) and (n x l) dimension.

An m x n matrix multiplied by an (n x l) vector results in an (m x l) vector.

\*When trying to get the prediction result, it is more efficient to use matrix multiplication than to build a for formula to calculate the result.

* + **Properties of matrix multiplication:**
    1. **It is not commutative:**

Let A and B be matrices. Then in general, A x B ≠ B x A.

* + 1. **It is associative:**

A x (B x C) = (A x B) x C

* + 1. Identity Matrix:

Denoted by I or

1s in the diagonal and 0s elsewhere.

A x I = I x A = A

* + **Matrix inverse:**

If A is an m x m matrix, and if it has an inverse, then .

Matrices that do not have inverse are called singular or degenerate.

* + **Matrix transpose:**

Let A be an m x n matrix, and let B = . Then B is an n x m matrix, and