Machine Learning Lab Class I and II

Ankit Satpute, 120825
Hsueh Wei, 120820
Sagar Nagaraj Simha, 120797
(M.Sc. CS4DM)

09/11/2019

Exercise 1.a

- Supervised Learning: Learns a function in a set of input output pairs. In this case the learning is carried with help of the supervision. We assume that at each instance of time when the input is applied, the desired response of the system is provided by the teacher and then difference between desired response and actual response (calculated) serves as an error measure which is used to correct parameters(weights).[4]
- Unsupervised Learning: Identify structures in data. This is learning without supervision i.e. desired response is unknown. In this type, learning is accomplished based on observations of responses to inputs that we have marginal to no knowledge about. Suitable weight self adaptation mechanism is needed in this case for error correction.[4]

Exercise 1.b

- Supervised Learning - Predicting the delay of an airline on a particular day of the week based on airline records.
- Unsupervised Learning - Grouping/Clustering of species according to similar genes.
- Reinforcement Learning - A learning system strategies to win the game of chess where the system figures out the right moves based on a reward.

Exercise 1.c

Considering an example of 'A checkers learning problem' of:
Task T : playing checkers.
Performance measure P: percent of games won in the tournament.
Training experience E: games played against itself.

A learning system has the following design choices to be made,

1. Choosing the Training Experience:
   (a) Direct feedback or Indirect feedback
   (b) The degree to which the learner controls the sequence of training examples
   (c) How well it represents the distribution of examples over the final system performance ‘P’ (Distribution of data over train and test samples is preferred to be similar).
2. Choosing the Target function
   (a) The type of knowledge that would be learned. Also called the Function Approximation. In case of the checkers problem, choosing a move from a set of legal moves.
3. Choosing the representation for the Target function
   (a) Example, quadratic polynomial function or an artificial neural network.
   (b) Reduces the problem of learning strategy to the problem of learning values of coefficients/weights.
4. Choosing a Function Approximation algorithm
   (a) The choice of the algorithm (Ex: Gradient Descent).
   (b) Estimating training values
   (c) Adjusting the weights (Ex: Minimizing the error between the training values and the approximated values)
5. The Final design
   (a) The Performance System
   (b) The Critic
   (c) The Generalizer
   (d) The Experiment Generator [1]

Exercise 1.d

This is a Supervised Learning paradigm.
Task ‘T’ - Predicting how high players can jump.
Performance measure ‘P’ - Number of players that jump the predicted height
Training Experience ‘E’ - A database of players with their height of jump

The model formation function in this learning system represents the player’s characteristics that contributes to the height of the jump (mapping to feature space). A useful model formation, 'alpha', that maps to feature space x = alpha(o), where 'o' is the player is, Height of the player, Weight of the player and Age of the player. The basketball trainer is the 'Y' (gamma) in this problem. 'O' is a set of players, provided that the data on height of players corresponding to their features is known to the basketball trainer

Exercise 1.e

1. One cannot learn about the alternate events that haven't happened and where data towards it does not exist. For example, suppose we have data about the lifespan of patients (with a condition) after administering a particular drug. From that, we cannot learn about the lifespan of the patients with the same condition who did not take the drug. A philosophical generalization on similar lines - One cannot learn about the future, for it hasn't happened yet or there exists no data about it. [2]

2. A problem cannot be solved by learning when there is no incentive to learn/solve a problem. Ex : Evolution may optimise survival of life forms only in case of hostile environment. A perfectly hospitable environment will not encourage learning for may be the system has reached saturation.

Exercise 2.a

First we calculate values required for the weight calculations based on the data given in the question. Obtained values are arranged in tabular form below:

x	y	(x-$\bar{x}$)	(y-$\bar{y}$)
5	50	-8.15	-88.25
7	79	-6.75	-59.25
15	124	1.25	-14.25
28	300	14.25	161.75

Formula for calculating weight w_0 is

$$w_0 = \bar{y} - \hat{w_1} * \bar{x}$$

Formula for calculating weight w_1 is

$$w_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) * (y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Now we substitute the values from above table in the formula given and calculate 2 weight values

$$w_1 = \frac{3459.25}{326.75} = 10.59$$

$$w_0 = 138.25 - (10.59) * 13.75 = -7.32$$

The weights w_0 and w_1 for the linear regression for the age variable are -7.32 and 10.59.

Exercise 2.b

Now we will extrapolate average stopping distance for 15 year old car i.e. just taking values of weights we calculated and substitute with age value for x in the equation

$$\bar{y} = -7.32 + 10.59 * x$$

$$\bar{y} = -7.32 + 10.59 * 15 = 151.53$$

After extrapolation the average stopping distance is 151.53

Exercise 2.c and 2.d

To avoid extensive calculations by hand we directly calculated the desired value with help of small python code which is attached in our code file. As a conclusion:
The weights after considering the mileage of the cars as additional variable are:
[1.28163977e+01] [-2.29034854e-04] [-6.42631756e+00]
Extrapolated average stopping distance for 15 year old car is: [149.19720329]

Exercise 2.e

Pitfalls and problems of the extrapolation [3].

1. In case of prediction outside the known range, the true function c(x) may perhaps be non linear or a different model beyond the data points to which the linear function (calculated) has been regressed to. In that case the extrapolation (*predicted value) would be vastly different from the true value and hence erroneous.

2. One can see from the figure of Anscombe's quartet dataset below that the distribution of data points and hence the underlying models in each case is different but the regressed line is the similar.

3. In case of regressed line y(x) with relatively higher error w.r.t the original c(x), the extrapolated values may have higher error too!

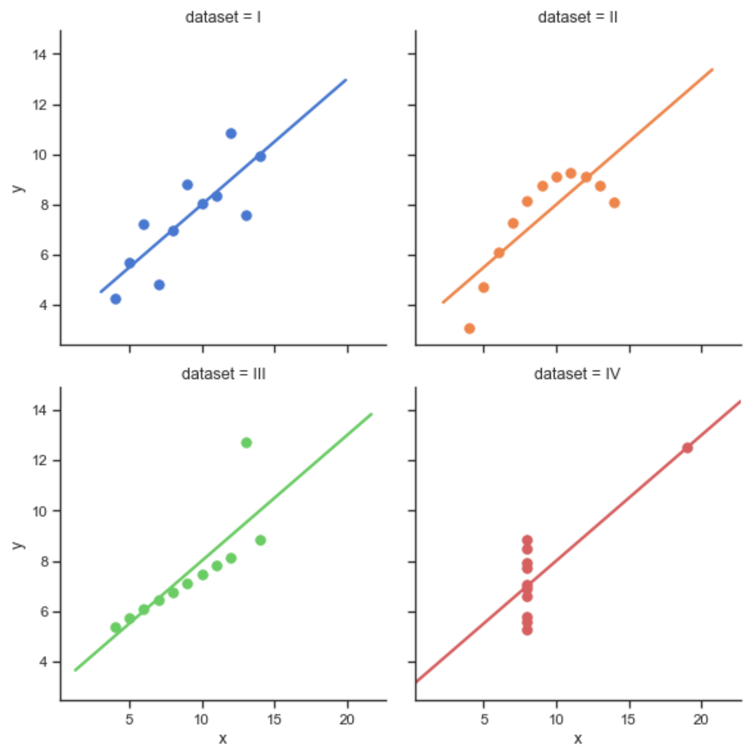

Fig: Anscombes Quartet

References

1 Tom Mitchell. Machine Learning. 1st edition, McGraw-Hill, 1997.

2 The Book of Why: The New Science of Cause and Effect, By Dana Mackenzie and Judea Pearl

3 https://stats.stackexchange.com/questions/219579/what-is-wrong-with-extrapolation

4 Jacek M. Zurada. Introduction to Artificial Neural System, West publishing company 1992.