COMP3010J Tutorial

Multivariate Linear Regression

1. Heating load is the amount of heat energy required to keep a building at a specified temperature, during the winter regardless of outside temperature. A multivariate linear regression model has been built to predict the heating load in a residential building based on a set of descriptive features describing the characteristics of the building. The descriptive features used are: the overall surface area of the building, the height of the building, the area of the building's roof, and the percentage of wall area in the building that is glazed. This kind of model would be useful to architects or engineers when designing a new building. The trained model is as follows:

HEATING LOAD = -26.030 + 0.0497 * SURFACE AREA
+ 4.942 * HEIGHT – 0.090 * ROOF AREA
+ 20.523 * GLAZING AREA

a) Use this model to make predictions about the query instances given below

ID	SURFACE AREA	HEIGHT	ROOF AREA	GLAZING AREA
1	784.0	3.5	220.5	0.25
2	710.5	3.0	210.5	0.10
3	563.5	7.0	122.5	0.40
4	637.0	6.0	147.0	0.60

b) The historical data collected for the query instances is given as below. Calculate the sum of squared errors and the `errorDelta` for your predictions.

ID	HEATING LOAD
1	22.7
2	9.8
3	37.2
4	39.0

2. A multivariate logistic regression model has been built to predict the propensity of shoppers to perform a repeat purchase of a free gift that they are given. The descriptive features used by the model are the age of the customer, the socio-economic band to which the customer belongs (a, b, or c), the average amount of money the customer spends on each visit to the shop, and the average number of visits the customer makes to the shop per week. This model is being used by the marketing department to determine who should be given the free gift. The weights in the trained model are shown in the table below

Feature	Weight
Intercept ($w[0]$)	-3.82398
AGE	-0.02990
SOCIO ECONOMIC BAND B	-0.09089
SOCIO ECONOMIC BAND C	-0.19558
SHOP VALUE	0.02999
SHOP FREQUENCY	0.74572

Using this model, make a prediction for each of the query instances given below. The following assumptions can be made:

   i.   The positive level is *yes*
  ii.   The classification threshold is 0.5

ID	AGE	SOCIO ECONOMIC BAND	SHOP FREQUENCY	SHOP VALUE
1	56	b	1.60	109.32
2	21	c	4.92	11.28
3	48	b	1.21	161.19
4	37	c	0.72	170.65
5	32	a	1.08	165.39