Practical Data Science/Analytics (Classification)

Write R Scripts or use R to perform any mathematical operations while solving the following problems.

Problem 1: KNN Learning

Given the following training data, predict the class of the following new example using k-Nearest Neighbour for k=5: age<=30, income=medium, student=yes, credit-rating=fair. For similarity measure use a simple match of attribute values:

$$\text{Similarity(A,B)} = \sum_{i=1}^{4} w_i * \partial(a_i, b_i) \Big/ 4 \quad \text{where } \partial(a_i, b_i)$$

is 1 if a_i equals b_i and 0 otherwise. a_i and b_i are either *age*, *income*, *student* or *credit_rating*. Weights are all 1 except for income it is 2.

RID	age	income	student	credit_rating	Class: buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31 . . . 40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31 . . . 40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31 . . . 40	medium	no	excellent	yes
13	31 . . . 40	high	yes	fair	yes
14	>40	medium	no	excellent	no

Practical Data Science/Analytics (Classification)

Problem 2: Naïve Bayes Learning (Only with Nominal Attributes)

Given the training data in Problem 1(Buy Computer data), predict the class of the following new example using Naïve Bayes Classification: age<=30, income=medium, student=yes, credit-rating=fair. Use Laplace's Correction factor while estimating likelihoods.

Problem 3: Naïve Bayes Learning (Nominal & Numerical Attributes)

Given the training data in the table below (*Tennis* data), predict the class of the following new example using Naïve Bayes classification: outlook=overcast, temperature=60, humidity=62, windy=false. Assume Gaussian distribution for numerical attributes and use Laplace's Correction factor while estimating likelihoods.

outlook	temperature	humidity	windy	play
sunny	85	85	false	no
sunny	80	90	true	no
overcast	83	86	false	yes
rainy	70	96	false	yes
rainy	68	80	false	yes
rainy	65	70	true	no
overcast	64	65	true	yes
sunny	72	95	false	no
sunny	69	70	false	yes
rainy	75	80	false	yes
sunny	75	70	true	yes
overcast	72	90	true	yes
overcast	81	75	false	yes
rainy	71	91	true	no

Problem 4: Decision Tree Learning

Given the training data in Problem 1(Buy Computer data), build a decision tree and predict the class of the following new example: age<=30, income=medium, student=yes, credit-rating=fair. Use Information-Gain as feature selection criteria while building decision tree.