CS6316 Exam – Fall 2017

SOLUTIONS

---(Exam Page 2)---

Question 1 [**10 points**]:

(a) `[4 pts]`

Let's assume a 'dumb' classifier classifies all instances to class 'A'. This will result in a misleading 90% accuracy (high). This is especially prevalent when the dataset is unbalanced. Any such explanation will be accepted.

(b) `[2 pts]`

Matthews Correlation Coefficient (MCC) – a measure of the quality of binary classifications. It is regarded as a balanced measure which can be used even if the classes are of very different sizes.

(c) `[4 pts]`

Stratified k-fold cross validation – maintains the proportion of the classes in each fold, so as to obtain a more accurate evaluation model (each fold is more balanced.)

[Each fold should have examples from all of the classes, and each fold should also have the same ratio of each of the classes.]

---(Exam Page 3)---

Question 2 [**10 points**]:

(a) `[-1 pts if no work shown]`

$-5/8 \log_2 5/8 - 3/8 \log_2 3/8 = \boxed{\mathbf{0.9544}}$ `[3 pts]`

(b) (i) **Attribute**: IsSmooth `[3 pts]`

(ii) **Information gain:** approximately **0.0487** `[4 pts]`

[0 = No; 1 = Yes.]

E(IsPoisonous, IsSmooth) = P(Yes)E(Yes) + P(No)E(No)

P(IsSmooth=Yes) = 4/8 = 0.5

$E(Yes) = -3/4 \log_2 3/4 - 1/4 \log_2 1/4 = 0.3112 + 0.5 = \mathbf{0.8112}$

IsPoisonous=Yes – IsPoisonout=No

P(IsSmooth=No) = 4/8 = 0.5

$E(No) = -2/4 \log_2 2/4 - 2/4 \log_2 2/4 = \mathbf{1}$

IsPoisonous=Yes – IsPoisonout=No

E(IsPoisonous, IsSmooth) = 0.9056

IG(IsPoisonous, IsMooth) = E(IsPoisonous) – E(IsPoisonous, IsSmooth)

$= 0.9546 - 0.9056 = \boxed{\mathbf{0.049}}$

---(Exam Page 4)---

Question 3 [10 points]:

Solution is to use Bayes Theorem: `[6 pts: using correct formula]`

$$P[B_j|A] = \frac{P[A \cap B_j]}{P[A]} = \frac{P[A|B_j]P[B_j]}{\sum_{k=1}^{N} P[A|B_k]P[B_k]}$$

The probability that the skillet was manufactured in factory B is: `[4 pts correct ans]`

$$\frac{(0.02 * 0.3)}{(0.01 * 0.5 \; + \; 0.02 \; * \; 0.3 \; + \; 0.03 \; * \; 0.2)} = \mathbf{0.353}$$

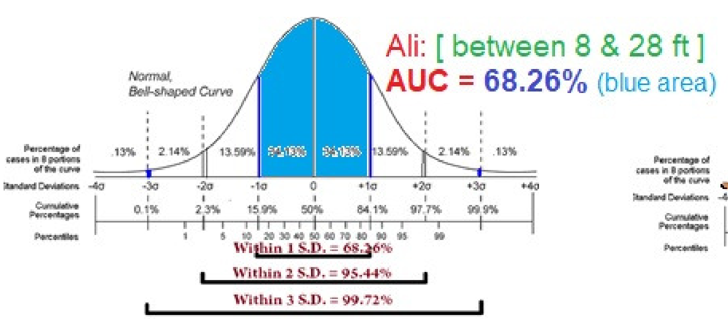
Question 4 [10 points]:

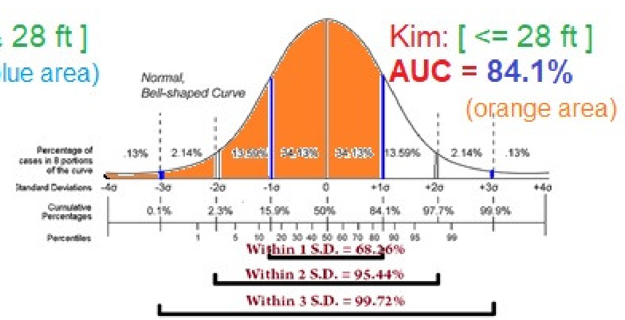
`[4 pts each for calculation of Kim and Ali's %, or explanation]`

Kim: The area under a normal curve between the left hand side of the curve and +1 standard deviation (28 feet) is 84.1%

(if 68.26% between +/- 1 st. dev., then area under left tail will be 15.87%) So 68.26 + 15.87=**84.1%**

Ali: The area under a normal curve between +/- 1 standard deviation is **68.26%** (8 to 28 feet)





Therefore, **Kim's prediction is most likely to be true** (84.1% > 68.26%) `[2 pts]`

---(Exam Page 5)---

Question 5 [10 points]: \

`[2 pts each  ~ 1pt for Yes/No, 1pt for explanation]`

(a) k-NN regression:
YES, because different scaling of inputs will affect the distance metric used in kNN algorithm

(b) Linear regression:
NO, pre-scaling is not needed for standard linear regression (as in classical statistics)

(c) SVM regression:
YES, because SVM is based on the notion of similarity (kernel) between input samples

(d) MLP regression:
YES, because the input layer mapping computes the distance (dot product) between multivariate input x and the vector of input weights w

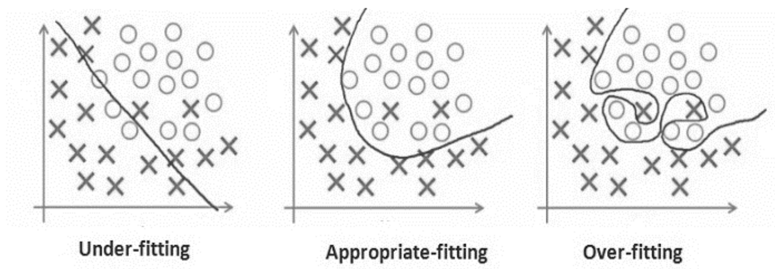
(e) Preprocessing task
Outlier removal – Most parametric statistics, like means, standard deviations, and correlations, and statistics based on these, are highly sensitive to outliers. Since the assumptions of common procedures, like linear regression, are also based on these statistics, outliers can really skew the analysis.

Many correct answers. Solution above is just one example.
Full credit as long as name and correct description of a preprocessing task is given.
No credit if task is not related to preprocessing.

---(Exam Page 6)---

Question 6 [10 points]:

In order to attain the best prediction accuracy the model must not overfit the data nor must it underfit the data. Also, according to **Occam's Razor**, the solution that is the **simplest** that does explain the data is the best. Student 1's model while simplest doesn't fit / explain the data (**underfitting**) so it can't be considered. Between Student 2's and Student 3's model, Student 2's model will attain the best prediction accuracy. It is the simplest compared to Student 3's (**overfitting**.)



[6 pts: discussion relating to Occam's Razor and clear discussion comparing the 3 students (using terms 'underfitting' and 'overfitting' is ideal).]
[4 pts: saying Student 2's model is best]

Question 7 [12 points]:

[4 pts each]

(a)

Scientific, because it can be tested and falsified. You can use a mirror, and send a probe to the moon!

(b)

Not scientific, because nobody knows what constitutes all factors → it cannot be tested. However, if **all such factors** are explicitly given *a priori*, then it is **scientific**. The statement in its present form does not specify all these factors, but presumes they exist.

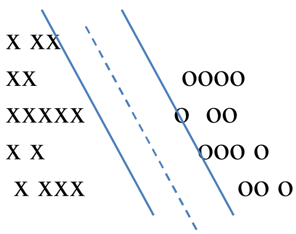
(c)

Scientific, because it can be tested and falsified (i.e. it is logically possible that there exist objects that do not require the same time to hit the ground)

---(**Exam Page 7**)---

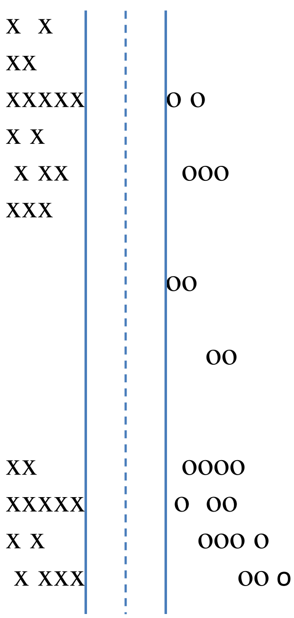
Question 8 [18 points]:

(a)



[6 points] ~ Orientation can be different

(b)



[6 points] ~ Generally no other orientation fits

(c)

Circle: [**Yes**] No [2 points] ~ but base this on student's (a)/(b)

(d)

[4 points] ~ base this generally on student's (a)-(c)

Even though the **margin size is the same**, linear SVM model (a) is more robust because it has a much **smaller range** of x-values than model (b). So Model (a) is expected to have smaller test error.

---(**Exam Page 8**)---

Question 9 [20 points]:

[(a) - (e): 4 pts each]

input x-axis → X X X OX O O O Items in **red** are misclassified (e.g. O)

For this data set:

(a)

X X X OX O O O **OR** X X X OX O O O

(b)

X X X OX O O O **OR** X X X OX O O O

(c)

X X X OX O O O

looking at nearest neighbors, so…
everything to the left of this line is closer to the 'x' class,
everything to the right of this line is closer to the 'o' class
assumption: second gap is larger, causing more separation

(d)

X X X OX O O O **OR** X X X OX O O O

(e)

X X X OX O O O

End of the Exam!
