COMP30027 Machine Learning
Practice Exam

Semester 1, 2019

Tim Baldwin, Afshin Rahimi, and Jeremy Nicholson

— version: 198, date: May 23, 2017 —

Lecture Outline

1. Practice Exam
   - Section A: Short Answer Questions
   - Section B: Method Questions
   - Section C: Numeric Questions
   - Section D: System Building Questions

Section A: Expectations

Answer each of the questions in this section as briefly as possible. Expect to answer each sub-question in a couple of lines at most.

Question 1: Short Answer Questions

a. Classify the following attributes as discrete, ordinal or continuous:
   (1) number of patients in a hospital

Question 1: Short Answer Questions

a. Classify the following attributes as discrete, ordinal or continuous:

(1) number of patients in a hospital

A: *[CONTINUOUS]*

Question 1: Short Answer Questions

a. Classify the following attributes as discrete, ordinal or continuous:

(1) number of patients in a hospital

A: *[CONTINUOUS]*

(2) ability to pass light, in terms of the following values: opaque, translucent and transparent

Question 1: Short Answer Questions

a. Classify the following attributes as discrete, ordinal or continuous:
   (1) number of patients in a hospital
   A: *[CONTINUOUS]*
   (2) ability to pass light, in terms of the following values: opaque, translucent and transparent
   A: *[ORDINAL]*

Question 1: Short Answer Questions

b. What is the difference between classification and regression?

Question 1: Short Answer Questions

b. What is the difference between classification and regression?

A: *In classification we predict a discrete class; in regression we predict a continuous numeric value.*

Question 1: Short Answer Questions

c. Outline the nature of "hill climbing" and provide an example of a hill climbing algorithm. What basic approach does hill climbing contrast with?

Question 1: Short Answer Questions

c. Outline the nature of “hill climbing” and provide an example of a hill climbing algorithm. What basic approach does hill climbing contrast with?

A: *Hill climbing is a greedy search technique where each “decision” is made based on the local criteria (usually the local optimum). An example of a hill climbing algorithm is ID3. Hill climbing contrasts with global/exhaustive search.*

Question 1: Short Answer Questions

d. Optionally with the use of a diagram, explain what it means for a dataset to be **linearly separable**?

Question 1: Short Answer Questions

d. Optionally with the use of a diagram, explain what it means for a dataset to be **linearly separable**?

A: *A linearly separable dataset is one where it is possible to partition the (training) instances perfectly based on class membership, using linear decision boundaries*

Question 1: Short Answer Questions

e. One advantage of **bagging** is that is can be **parallelised**: how and why?

Question 1: Short Answer Questions

e. One advantage of **bagging** is that is can be **parallelised**: how and why?

A: *In each iteration of bagging, the algorithm independently samples from the training data, and independently trains a model over that sample. Because of the complete independence of the individual iterations, it is possible to train the base models in parallel.*

Section B: Expectations

In this section you are asked to demonstrate your conceptual understanding of a subset of the methods that we have studied in this subject.

Question 2: Logistic Regression

For a **logistic regression** model trained over n classes, based on k training instances and f features, and tested over t test instances, use big-O notation to describe the approximate number of parameters in the trained model.

Question 2: Logistic Regression

For a **logistic regression** model trained over n classes, based on k training instances and f features, and tested over t test instances, use big-O notation to describe the approximate number of parameters in the trained model.

A: $\mathcal{O}(nf)$

Question 3: Evaluation

a. How does stratified cross validation contrast with (non-stratified) cross validation?

Question 3: Evaluation

a. How does stratified cross validation contrast with (non-stratified) cross validation?

   A: *In stratified cross-validation, the class distribution is balanced in each partition of the data, whereas in (non-stratified) cross validation, there is no such guarantee.*

Question 3: Evaluation

b. What is "sum of squared errors"? Name a learning task it is applied to.

Question 3: Evaluation

b. What is "sum of squared errors"? Name a learning task it is applied to.

A: *Sum of squared errors (SSE) is used in clustering evaluation, e.g., to calculate the square of the distance between each instance and its cluster centroid*

Section C: Expectations

In this section you are asked to demonstrate your understanding of a subset of the methods that we have studied in being able to perform numeric calculations over a given set of instances.

Question 5: Naive Bayes

1. Estimate the conditional probabilities for $P(A|+)$, $P(B|+)$ and $P(C|-)$ based on the following training dataset *without* smoothing:

A	B	C	Class
0	0	0	$+$
0	0	1	$-$
0	1	1	$-$
0	1	1	$-$
0	0	1	$+$

A	B	C	Class
1	0	1	$+$
1	0	1	$-$
1	0	1	$-$
1	1	1	$+$
1	0	1	$+$

A:

$$\begin{aligned} P(A=1|+) &= \frac{3}{5} \quad (=0.6) \\ P(B=1|+) &= \frac{1}{5} \quad (=0.2) \\ P(C=1|-) &= \frac{5}{5} \quad (=1.0) \end{aligned}$$

Question 5: Naive Bayes

2. Predict the class label for $(A = 0, B = 1, C = 0)$ using the given training dataset and non-smoothed probability estimates, based on the naive Bayes approach.

   A:

$$\begin{aligned} Class\ +: \quad & \tfrac{1}{2}(\tfrac{2}{5} \times \tfrac{1}{5} \times \tfrac{1}{5}) &=& \quad \tfrac{1}{125} \quad (= 0.008) \\ Class\ -: \quad & \tfrac{1}{2}(\tfrac{3}{5} \times \tfrac{2}{5} \times \tfrac{0}{5}) &=& \quad 0 \end{aligned}$$

$$Most\ probable\ class\ c = +$$

Question 5: Naive Bayes

3. Estimate the conditional probabilities for $P(A|+)$, $P(B|+)$ and $P(C|-)$ based on the given training dataset with Laplacian smoothing.

A:

$$\begin{aligned} P(A=1|+) &= \frac{4}{7} \ (\approx 0.57) \\ P(B=1|+) &= \frac{2}{7} \ (\approx 0.29) \\ P(C=1|-) &= \frac{6}{7} \ (\approx 0.86) \end{aligned}$$

Question 5: Naive Bayes

4. Predict the class label for $(A = 0, B = 1, C = 0)$ using the given training dataset and smoothed probability estimates using Laplacian smoothing, based on the naive Bayes approach.

   A:

$$
\begin{aligned}
Class\ + : \quad & \tfrac{1}{2}\left(\tfrac{3}{7} \times \tfrac{2}{7} \times \tfrac{2}{7}\right) &=& \tfrac{6}{343} \quad (\approx 0.017) \\
Class\ - : \quad & \tfrac{1}{2}\left(\tfrac{4}{7} \times \tfrac{3}{7} \times \tfrac{1}{7}\right) &=& \tfrac{6}{343} \quad (\approx 0.017)
\end{aligned}
$$

$$Most\ probable\ class\ \mathsf{c} = ?$$

Question 5: Naive Bayes

5. With reference to the underlying assumptions made in Laplacian smoothing, explain the observed differences in the predictions from Parts 2. and 4.

   A: *Laplacian smoothing assumes that every event has non-zero likelihood, which it derives by assigning a (hidden) training instance to each combination of feature value and class. This prevents the unobserved combination of* $-$ *and* $C = 0$ *from decimating all other probabilities in 2. With smoothing in (d), both* $+$ *and* $-$ *have non-zero probability estimates, and are ultimately tied. Possibly this is an overestimate of the likelihood of class* $-$*, due to the heavy-handedness of this assumption.*

Section D: Expectations

In this section you are asked to demonstrate that you have gained a high-level understanding of the methods and algorithms covered in this subject, and can apply that understanding. Expect to respond to each question using about one-third to one-half a page. These questions will require significantly more thought than those in Sections A–C and should be attempted only after having completed the earlier sections.