Q1
Consider the following data set:

a1	*a2*	*a3*	*Class*
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	1	0	0
1	1	1	0

What values would the Naive Bayes classifier assign to p(Class=0) and p(Class=1) if no smoothing is used?

Ans1
p(Class=0) = 4/7
p(Class=1) = 3/7

Q2
Consider the following data set:

a1	*a2*	*a3*	*Class*
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	1	0	0
1	1	1	0

What is the value of p(a1=1|Class=0)?

Ans2
1/2

Q3
Consider the following data set:

a1	*a2*	*a3*	*Class*
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	1	0	0
1	1	1	0

What is the value of p(Class=0|a1=1)?

Ans3
2/3

Q4
Which validation strategy would you recommend to estimate the performance of k-nearest neighbors on a large dataset? We want the best estimate of the performance that can be computed in a reasonable amount of time.

Ans4
Splitting data into training and test set

Q5
A binary classifier produced the following confusion matrix:

	Pred = 0	*Pred = 1*
Class = 0	80	5
Class = 1	7	8

What is the classifier's accuracy?
Ans5
88/100

Q6
A binary classifier produced the following confusion matrix:

	Pred = 0	*Pred = 1*
Class = 0	80	5
Class = 1	7	8

What is the classifier's precision?
Ans6
8/13

Q7
A binary classifier produced the following confusion matrix:

	Pred = 0	*Pred = 1*
Class = 0	80	5
Class = 1	7	8

What is the classifier's recall?
Ans7
8/15

Q8
A binary classifier produced the following confusion matrix:

	Pred = 0	*Pred = 1*
Class = 0	80	5
Class = 1	7	8

What is the classifier's f-measure?
Ans8
0.5714

Q9
Let C be an unweighted k-nearest neighbors classifier where k is equal to the size of the training set. C will predict the majority class in the training set for all examples in the test set.
Ans9
True

Q10
A weighted k-nn classifer with k=1 will make exactly the same predictions as a weighted k-nn classifer with k=2.
Ans10
True

Q11
A weighted k-nn classifer with k=1 will make exactly the same predictions as an unweighted k-nn classifer with k=1.
Ans11
True

Q12
Suppose we train a classifier and evaluate its performance using the same dataset (i.e. we don't use separate training and test sets). This results in an overoptimistic performance prediction. Which algorithm will have its performance most over-estimated? (i.e. the difference between the estimated performance and the performance on unseen data will be largest).
Ans12
k-nearest neighbors

Q13
Suppose we train a classifier and evaluate its performance using the same dataset (i.e. we don't use separate training and test sets). This results in an overoptimistic performance prediction. Which algorithm will have its performance least over-estimated? (i.e. the difference between the estimated performance and the performance on unseen data will be smallest).
Ans13
Linear regression

Q14
Linear regression cannot detect non-linear relationships among attributes. What can be done to reduce the impact of this problem?
Ans14
We can add non-linear attributes by appending non-linear properties such as the cube of variables and square of variables. to features.

Q15
Results provided by k-nearest neighbors often depend on the units in which attributes are measured. What can be done to reduce the effect of this limitation?
Ans15
Scaling can be used in order to normalize or standardize the features. In normalization we scale the attribute values to the 0-1 range, in normalization, we transform the attributes to have zero meand and unit variance.

Q16
Naive Bayes computes predictions by multiplying an often large number
of conditional probabilities. This often results in numeric underflow (i.e. the product becomes 0). What can be done to reduce the effect of this limitation?
Ans16
To fix this problem, you can modify the equation so that instead of multiplying by each probability, you are adding the log of it.