CS 361 Machine Learning Exam 1 (January-May Session, 2021)

Total No.: 20 Attempt all questions Time: 30 Minutes

Points: 14/20

1

Which of the following sentence is FALSE regarding regression? (1/1 Point)

- [ ] It relates inputs to outputs
- [ ] It is used for prediction
- [ ] It may be used for interpretation
- [x] It discovers causal relationships ✓

2

In the case of supervised learning, which statement(s) is/are NOT TRUE? (1/1 Point)

- ( ) It prefer the simplest hypothesis consistent with data
- ( ) Given a collection of examples of f, return a function h that approximates f.

- [x] Feature space is always same as input space ✓
- [ ] Can be called as Inductive Learning

3

Which of the following is/are true about Maximum Likelihood Estimate (MLE)?
(1/1 Point)

- [x] MLE does not necessarily exist. ✓
- [ ] MLE always exists
- [x] If exists/exist, it/they may not be unique ✓
- [ ] If exists/exist, it/they must be unique

4

102.9	16.7
109.9	18.7
116.1	20.7
121.7	22.9
127	25.3
132.2	28.1
137.5	31.4
140	32.2
147	37
153	40.9
160	47
166	52.6

Consider the given figure that shows a sample data consisting of two features X1 and X2, and a class label Y. Assume that k-NN is used for memorizing this data and performing classification on some test points. But observing the data and labels, there could be some problems if k-NN memorizes this given data. What operations can be done so that k-NN can generally work well with such data too?
(0/1 Point)

- [ ] Transform the data by column-wise subtracting the mean and dividing by the standard deviation ✓
- [ ] Multiply class of each of the K neighbors by a weight proportional to the inverse of the distance from that training point to the given test point ✓
- [x] Plotting a graph of validation error-rate vs. k-value and get the best value of k
- [x] Checking different distance metrics and deciding which performs the best

✗

5

In the context of "Curse of Dimensionality", which of the following statement(s) is/are TRUE?
(0/1 Point)

- [x] The distance between two randomly drawn data points increases drastically with their dimensionality. ✓
- [x] One possible solution is to increase the number of training samples, n, until the nearest neighbors are truly close to the test point.
- [x] In d dimensions, d−1 dimensions will be orthogonal to the normal of any given hyper-plane. ✓
- [x] The true dimensionality of the data may be much lower than its ambient space. ✓

6

Assertion (P): The path taken by Stochastic Gradient Descent (SGD) towards the minima has high variance and convergence is complicated.
Reason (R): In each iteration of SGD, random samples are picked based on which derivatives are calculated.
(1/1 Point)

- [x] Both P and R are true and R is the correct explanation of P ✓
- [ ] Both P and R are true but R is not the correct explanation of P
- [ ] Both P and R are false
- [ ] P is false but R is true

- [ ] P is true but R is false

7

MLE can be considered as minimizing the dissimilarities between p_data and p_model, where the dissimilarity is measured by using
(0/1 Point)

- [ ] MAP of p_data and p_model
- [x] Argmax of gaussian distribution estimators
- [x] KL Divergence ✓
- [x] Log-likelihood of Bayes error between p_data, and p_model

8

In the given figure, for a sample data distribution, X-axis and Y-axis denote the independent and dependent variables respectively. Which of the following are true if we try to model this distribution?
(1/1 Point)

- [ ] Simple linear regression will have low bias and high variance
- [x] Learning a quadratic fit will have low bias and low variance. ✓

- [x] Learning a polynomial fit of degree 3 will tend to overfit ✓
- [ ] Simple Linear regression will have high bias and high variance

9

In the context of regularization, which of the following is/are true?
(1/1 Point)

- [ ] Weights and biases are regularized using the same lambda
- [ ] Only the weights related to affine transformations are regularized in addition to biases
- [x] Biases are not regularized as un-regularized biases do not impose high variance ✓
- [ ] Weights and biases are regularized using the different lambda

✗

10

What is the use of a regularization parameter while performing a regularized linear regression?
(0/1 Point)

- [x] Controls the trade-off between the need for the model to fit the training set well and also have a large number of model parameters
- [ ] It reduces the bias in the model and hence reduces overfitting
- [x] Until some point, increasing it reduces the variance of the model significantly without significant addition of bias to the model ✓
- [ ] None of the above

✗

11

Which of the following statement(s) is/are TRUE?
(0/1 Point)

- [x] If the samples are drawn i.i.d. from the same distribution P, then the testing loss is an unbiased estimator of the true generalization loss. ✓
- [ ] We are trying to find a hypothesis h which would have performed well on the past/known data ✓
- [x] We evaluate our classifier on the testing loss ✓
- [ ] None of the above

12

Which of the following statement is/are NOT true for stochastic gradient descent?
(1/1 Point)

- [ ] In SGD, a few samples are selected randomly instead of the whole data set for each iteration
- [x] SGD is slower than Batch Gradient Descent ✓
- [ ] For SGD, path for reaching global/local minima is more randomized.
- [x] Normal gradient descent is less expensive (computation cost) than SGD. ✓

✗

13

Which of the following are true with respect to the Naive Bayes classifier?
(0/1 Point)

- [ ] Assumption of independent predictor variables ✓
- [ ] Should only be used when we have categorical input variables
- [x] It will make the correct MAP decision rule classification as long as the correct class is more probable than any other class ✓
- [x] We can consider Naive Bayes to be a log linear model ✓

14

Large values of the log-likelihood statistic indicate that
(1/1 Point)

- [ ] the statistical model fits the data well
- [ ] as the predictor variable increases, the likelihood of the outcome occurring decreases
- [ ] there are a greater number of explained vs. unexplained observations
- [x] statistical model is a poor fit of the data ✓

15

Consider a univariate linear regression model. Which of the following are true?
(1/1 Point)

- [ ] Changing the input variable by 1 unit always affects the output by 1 unit too
- [ ] Considering Mean Squared Error to compute the loss is a good idea as it reduces the effect of outliers
- [ ] Since it is univariate, we need to estimate one coefficient for modelling the data
- [ ] The decision boundary is a property of the data set given to us
- [x] None of the above ✓

16

If the cost function of Linear Regression is used for Logistic regression
(1/1 Point)

- [ ] Regularization for logistic regression may be hampered.
- [ ] Hypothesis function becomes intractable
- [x] Cost function will be non convex in nature ✓

- [ ] Cost function becomes logarithmic in nature

17

Which of the following about kNN classifiers are true?
(1/1 Point)

- [ ] Considering more neighbors for assessment of class votes improves classification accuracy.
- [x] There is no explicit training phase ✓
- [ ] Decision boundary is smoother with smaller values of k
- [x] Can be used for both classification and regression ✓

18

K-fold cross-validation is
(1/1 Point)

- [x] Linear in K ✓
- [ ] Quadratic in K
- [ ] Cubic in K
- [ ] Exponential in K

19

Bayesian networks allow the compact specification of
(1/1 Point)

- [ ] Conditional independence
- [ ] Belief
- [x] Joint probability distributions ✓
- [ ] Propositional logic statements

20

$$\frac{1}{2N}\sum_{i=1}^{N}(y_i - \beta_0)^2$$

For the one-parameter model, mean-Square error (MSE) is defined as given. We have a half-term in the front because,
(1/1 Point)

- [ ] scaling MSE by half makes gradient descent converge faster
- [ ] presence of half makes it easy to do grid search
- [x] it does not matter whether half is there or not ✓
- [ ] none of the above