MLBP 2019 - Quiz 1

13/09/2019 - 22/09/2019 at 23:59

---

(1) (1.0 pt.) Consider a dataset of 100 examples with exactly 40% noisy examples. What is the probability of having more than of half the noisy examples, in a sample of 50 drawn uniformly from the dataset.
Hint 1: you need to compute probability of having more than of 25 noisy examples in the drawn sample of size 50
Hint 2: You will be needed to sum up Binomial trials, where each example drawn for the sample can be either from the noisy examples or not
Hint 3: You can compute it by hand, but it could be faster computationally
1- 0.06
2- 0.15
3. 0.10

(2) (0.5 pt.) Based on PAC theory for the hypothesis class of axis parallel rectangles, the lower bound of the sample size ($T(\epsilon, \delta) = 4/\epsilon \log 4/\delta$) has a stronger dependence on:
1- error (ϵ)
2- confidence level (1- δ)
3- both the same effect

(3) (0.5 pt.) Which option is <u>not</u> considered as an assumption in PAC theory:
1- the data are sample from a uniform distribution
2- the underlying distribution is fixed but unknown
3- the examples should be independently drawn from an identical distribution

(4) (0.5 pt.) The expected generalization error with respect to generalization error bound, usually has:
1- higher value
2- lower value
3- the same value

(5) (1.0 pt.) Based on the generalization bound relying on the size of the hypothesis class using boolean conjunctions, and the following information, what is the lower bound on the number of examples:
(Formula: $m \geq \frac{1}{\epsilon}(\log(|\mathcal{H}|) + \log(\frac{1}{\delta})))$
Dataset : 3 binary features and one binary label
error bound : 10%
confidence level : 95% ($\delta = 5\%$)
1- 86
2- 157
3- 63

(6) (1.0 pt.) Based on the generalization bound for true error, using the empirical error : $R(h) \leq \hat{R}(h) + \sqrt{\frac{log\frac{2}{\delta}}{2m}}$, if we change δ, from 0.1 to 0.2, how many example will be needed to keep the same bound as before.
1- 0.5m
2- 1.6m

3- 0.8m

(7) (0.5 pt.) Which option is <u>not</u> correct based on the Bayes error

1- there might be a learner with a lower error than the Bayes error

2- Bayes error can be reduced by reducing the overlap between class distributions in the input space

3- Bayes error gives the expected noise level

(8) (2.0 pt.) What is the vector of coefficients for linear regression model, for the given inputs (x) and outputs (y). [computational work]

x = [0, 1, 3, 4, 6, 8, 9]

y = [2, 5, 7, 6, 8, 14, 12]

1- [2.70 1.13]

2- [2.01 1.56]

3- [1.59 1.96]

(9) (1.5 pt.) For the above information, which option is the result of $\frac{covariance(x,y)}{variance(x)}$. [computational work]

covariance: $cov_{x,y} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{N}$

variance: $var_x = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}$

1- 1.59

2- 2.01

3- 1.13

(10) (1.5 pt.) What is the predicted outputs for input test vector [2,7], based on the built LR model in question 8. [computational work]

1- [6.13, 13.72]

2- [4.97, 10.62]

3- [5.01, 11.25]