Machine Learning

Quiz & Notes on final exam

Quiz, 1

- (T or F) A classifier trained on $\boxed{\text{less training data}}$ is $\boxed{\text{less likely}}$ to $\boxed{\text{overfit.}}$

- Answer: **F**

Quiz, 3

- (T or F) If today I want to predict the $\boxed{\text{probability}}$ that a student sleep more than 8 hours on average (SA) given the Course loading (C), I will choose to use linear regression over logistic regression.
- Answer: **F**

Quiz, 8

Underfitting

- You train a linear classifier on 10,000 training points and discover that the training accuracy is only 67%. Which of the following, done in isolation, has a good chance of improving your training accuracy?

  A. Add novel features

  B. Train on more data

  C. Use linear regression

  D. Train on less data

- Answer: **AD**

Quiz, 9

Overfitting

- You train a classifier on 10,000 training points and obtain a training accuracy of 99%. However, when you submit to Kaggle (A testing platform), your accuracy is only 67%. Which of the following, done in isolation, has a good chance of improving your performance on Kaggle?

  A. Set your regularization value (λ) to 0

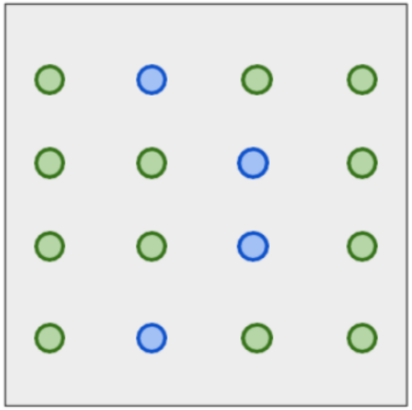
  B. Train on more data

  C. Use validation to tune your hyperparameters

  D. Train on less data

- Answer: **BC**

Quiz, 10



- You are performing classification in d dimensions (where d is very large) using a k-nearest neighbor classifier. You have N training samples and get an acceptable performance. Your boss hands you a similar classification task but in $2d$ dimensions. How many training samples will you need to get a comparable performance? Make an educated guess.

A. $2N$	B. e^N	C. N^2	D. $d^{0.5}$	E. e^d
F. $N^{0.5}$	G. $\log(N)$	H. d^2	I. $2d$	J. $\log(d)$

- Answer: **C** (KNN P37 Curse of dimensionality)

Quiz, 11

- Consider the following joint distribution on X and Y , where X∈{-1, 0, 1} and Y∈{0, 1}: p(X = -1, Y = 0) = 0.05, p(X = -1, Y = 1) = 0.05, p(X = 0, Y = 0) = 0.1, p(X = 0, Y = 1) = 0.1, p(X = 1, Y = 0) = 0.3, p(X = 1, Y = 1) =0.4.You learn that X ≥ 0. What is the largest probability of being correct you can achieve when predicting Y in this case?

A. 5/9 B. 1 C. 2/3 D. 1/4 E. 1/2
F. 1/7 G. 1/3 H. 6/7 I. 4/7 J. 3/7

- Answer: **5/9**

X \ Y	**0**	**1**
-1	0.05	0.05
0	0.1	0.1
1	0.3	0.4

$$P(Y = 1|X \geq 0) = \frac{P(X \geq 0, Y = 1)}{P(X \geq 0)}$$

Quiz, 13

- Consider a $\boxed{\text{linear regression}}$ problem with N samples $\{(x_n, y_n)\}(n = 1\ldots N)$, where each input x_n is a D-dimensional vector $\{-1, +1\}^D$, and all output values are $y_i \in R$. Which of the following statements is correct?

  A. Linear regression always "works" very well for N<<D

  B. A linear regressor works very well if the data is linearly separable.

  C. Linear regression always "works" very well for D << N

  D. None of the above.

- Answer: **D**

Quiz, 14A 16D

- (14) Which of the following are true about decision trees?

  A. They can be used only for classification.
- (16) Which of the following are true for the k-nearest neighbor (k-NN) algorithm?

  D. k-nearest-neighbors cannot be used for regression.
- Both choices are **wrong**.

Quiz, 21

- Consider the problem of binary classification using the Naive Bayes classifier. You are given two dimensional features (X_1, X_2) and the categorical class conditional distributions in the tables below. The entries in the tables correspond to $P(X_1 = x_1|C_i)$ and $P(X_2 = x_2|C_i)$ respectively. $\boxed{\text{The two classes are equally likely.}}$

Class / $X_1 =$	C_1	C_2
-1	0.2	0.3
0	0.4	0.6
1	0.4	0.1

Class / $X_2 =$	C_1	C_2
-1	0.4	0.1
0	0.5	0.3
1	0.1	0.6

- Given a data point $(-1, 1)$, calculate the following posterior probabilities:
  - $P(C_1|X_1\text{=-}1, X_2\text{=}1)$, $P(C_2|X_1\text{=-}1, X_2\text{=}1)$

Quiz, 21

Class / $X_1 =$	C_1	C_2
-1	0.2	0.3
0	0.4	0.6
1	0.4	0.1

Class / $X_2 =$	C_1	C_2
-1	0.4	0.1
0	0.5	0.3
1	0.1	0.6

- Given a data point $(-1, 1)$, calculate the following posterior probabilities:
  - $P(C_1|X_1\text{=-}1, X_2\text{=}1)$, $P(C_2|X_1\text{=-}1, X_2\text{=}1)$

$$\frac{P(X_1 = -1, X_2 = 1|C_1)P(C_1)}{P(X_1 = -1, X_2 = 1)}$$

$$= \frac{P(X_1 = -1|C_1)P(X_2 = 1|C_1)P(C_1)}{P(X_1 = -1|C_1)P(X_2 = 1|C_1)P(C_1) + P(X_1 = -1|C_2)P(X_2 = 1|C_2)P(C_2)}$$

$$= 0.1$$

Notes on final exam

- 不定项选择题：不建议选择没有把握的选项
- 注意题目中的信息：true, false, wrong, correct, incorrect
- 不能使用计算器，请注意计算准确度