University Jean Monnet   MLDM Master Program

"Introduction to Machine Learning" Exam (November 2023)

Guidelines:
- In the *Answer Sheet*, circle the letter (only one) corresponding to the correct answer.
- Each correct answer **adds one point**. Each incorrect answer **subtracts half a point**.

1. The height of a person is:

    a. An observable feature.

    b. A latent feature.

    c. A handcrafted feature.

2. Which of the following statements about **representation learning** is incorrect?

    a. It is computationally convenient at test time.

    b. It aims at learning latent features.

    c. It aims at learning handcrafted features.

3. Which of the following loss functions is not adapted to a classification task?

    a. $\ell(h, z = (x, y)) = e^{-yh(x)}$.

    b. $\ell(h, z = (x, y)) = (y - h(x))^2$.

    c. $\ell(h, z = (x, y)) = [yh(x) < 0]$, where $[.]$ is an indicator function.

4. Which loss function can be subject to gradient explosion?

    a. Hinge loss.

    b. Exponential loss.

    c. (0-1)-loss.

5. A classifier h_θ is obtained from S by solving the following optimization problem:

$$\min_\theta \sum_{i=1}^{n} \ell(h_\theta, S) + \lambda \|\theta\|_p$$

h_θ has a training error=2% and a test error=15%. What do you suggest?

    a. Decrease the number of training examples n?

    b. Use another loss function ℓ.

    c. Increase the value of the regularization parameter λ.

6. Which of the following statements about the regularization parameter λ is correct?

    a. A too large λ can cause your hypothesis to underfit the data.

    b. A too large λ can cause your hypothesis to overfit the data.

    c. Using a large value of λ cannot hurt the performance of your hypothesis.

7. Which ℓ_p-norm is convex?

    a. ℓ_0

    b. $\ell_{0.5}$

    c. ℓ_1.

8. The use of the ℓ_1-norm as a regularization term allows to:

    a. get a sparse model only.

    b. reduce the risk of overfitting only.

    c. both get a sparse model and reduce the risk of overfitting. ·

9. In a binary classification setting, an error occurs when:

    a. $yh(x) \neq 0$.

    b. $y - h(x) < 0$.

    c. $yh(x) < 0$.

10. Given the objective function $J(\theta) = \theta^2$. Given an initialization $\theta = -1$ and a learning rate $\alpha = 0.3$, what is the updated value of θ after the first iteration of gradient descent?

    a. $\theta = -1.6$. ·

    b. $\theta = -0.4$.

    c. $\theta = 0.8$.

11. As the number of features grows, the amount of data to learn well grows:

    a. linearly fast.

    b. quadratically fast.

    c. exponentially fast. ·

12. In the following figure, which hypothesis would be selected according to Occam's razor principle?
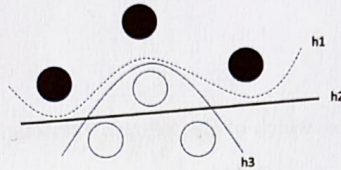
a. h_1.

b. h_2.

c. h_3.

13. What does the term PAC mean in statistical learning theory?

a. Probably Almost Correct.

b. Probably Approximately Correct.

c. Potentially Approximately Correct.

14. The optimal classifier h^* is usually not selected by the learning algorithm because:

a. It does not depend on the training examples.

b. It changes according to the algorithm.

c. It is not in the family of hypotheses.

15. Let $f_A(x) = x$, $\forall x \in [0, \sqrt{2}]$ and $f_B(x) = 1$, $\forall x \in [0, 1]$ be the distributions of the class A and B, respectively. What is the value of the bayesian error ϵ_B?

a. $\epsilon_B = \frac{1}{4}$

b. $\epsilon_B = \frac{1}{2}$

c. $\epsilon_B = \frac{2}{3}$

16. A training set S of 4 points is shattered by a family of hypotheses $\mathcal{H}$ if $\mathcal{H}$ realizes:

a. the 4 dichotomies of S.

b. the 8 dichotomies of S.

c. the 16 dichotomies of S.

17. Which of the following criteria is not a complexity measure?

a. The VC-dimension.

b. The number of training examples.

c. The Rademacher complexity.

3

18. The error obtained using K-cross-validation is an estimate of the generalization error. Is this estimate unbiased?

    a. Yes ·

    b. No

    c. It depends on K

19. Regarding bias and variance, which of the following statements is right?

    a. Models which overfit have a high bias.

    b. Models which overfit have a low bias. '

    c. Models which underfit have a high variance.

20. What is the Hessian matrix H of the function $f(x, y) = -xy + 2x^2y^4 - 3y^3$?

    a. $H = \begin{pmatrix} 4x^2 & 16xy^3 \\ 16y^3 & 12x^2y^2 - y \end{pmatrix}$

    b. $H = \begin{pmatrix} 4x^2 & 10xy^3 - 1 \\ 10xy^3 - 1 & 12x^2y^2 - y \end{pmatrix}$

    c. $H = \begin{pmatrix} 4y^{4}\text{\textbf{★}} & 16xy^3 - 1 \\ 16xy^3 - 1 & 24x^2y^2 - 18y \end{pmatrix}$

21. What is the incorrect statement?

    a. a k-NN approximates the bayesian classifier

    b. a k-NN learns spheres. ·

    c. a k-NN is lazy.

22. With noisy data, which of the following options would you consider for a k-NN classifier?

    a. Increase the value of k ·

    b. Decrease the value of k

    c. Dealing with noise is independent from the value of k

23. What happens when running a k-NN algorithm on imbalanced datasets composed of a few positives and many negatives?

    a. The decision boundaries are overly in favor of the positive class.

    b. The decision boundaries are not affected by the imbalance.

    c. The decision boundaries are overly in favor of the negative class.

24. With a kNN classifier, under which conditions $\frac{k_n}{nV_n}$ is a good estimate of $p(\mathbf{x})$, where V_n is a volume centered at x, k_n is the number of examples in V_n and n is the size of the training set?

   a. $\lim_{n \to \infty} V_n = 0$, $\lim_{n \to \infty} k_n = \infty$ and $\lim_{n \to \infty} \frac{k_n}{n} = 0$.

   b. $\lim_{n \to \infty} V_n = 0$, $\lim_{n \to \infty} k_n = 0$ and $\lim_{n \to \infty} \frac{k_n}{n} = 0$.

   c. $\lim_{n \to \infty} V_n = 0$, $\lim_{n \to \infty} k_n = 0$ and $\lim_{n \to \infty} \frac{k_n}{n} = \infty$.

25. Which statement is wrong?

   a. $\epsilon_{1NN} \le \epsilon_B$

   b. $\epsilon_B \le 2\epsilon_{1NN}$.

   c. $\epsilon_{1NN} \le 2\epsilon_B$

26. A training set is composed of three examples $(x_1, x_2) \in \mathbb{R}^2$, where the label $y \in \{-1, +1\}$.

x_1	x_2	y
a	c	+1
b	d	+1
a	b	-1

Using the naive Bayes classifier, what is the value of $P(y = +1 | x_1 = a, x_2 = d)$?

   a. 0.

   b. $\frac{1}{3}$.

   c. $\frac{1}{2}$.

27. Which of the following statements about the naive Bayes classifier is wrong?

   a. It overcomes the problem of curse of dimensionality.

   b. It assumes that the features are not independent.

   c. It allows to smooth the probabilities.

28. The Bayesian error is supposed to be $\epsilon_B = 0\%$. A classifier h has a training error=9% and a validation error=10%. What do you suggest?

   a. To increase the number of training examples.

   b. To increase the complexity of your classifier.

   c. To reduce the complexity of your classifier.

29. Given the following ranking obtained from 3 positive examples (in red) and 5 negatives (in blue). What is the value of the Average Precision AP?
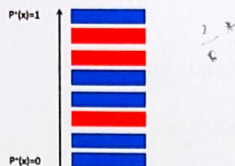
a. $\frac{5}{9}$.

b. $\frac{2}{11}$.

c. $\frac{2}{7}$.

30. Which of the following is incorrect about the F_β-measure?

a. The F_β-measure is not convex.

b. The F_β-measure is defined as follows: $F_\beta = \frac{(1+\beta^2)Precision+Recall}{\beta^2 \times Precision \times Recall}$

c. F_β is a non separable function. ·