Université Nice Sophia-Antipolis
Statistical leaarning, 2021-2022
`https://math.unice.fr/~rubentha/cours.html`

Final examination (IM), A

Documents and calculators forbidden.Give back the subject with your copy (+0.5 points!).).
Duration: 2h30.

Part 1. Multiple choice questions (10 points, write the answers on the examination copy, without justification (this is a quiz). One answer per question, one point for a correct answer (zero point otherwise))

(1) You are working on a project that involves clustering together images of different dogs. You take image and identify it as your centroid image. What type machine learning algorithm are you using?
  (a) Centroid reinforcement.
  (b) Binary classification.
  (c) K-means clustering.
  (d) K-nearest neighbour.

(2) What is overfitting?
  (a) When a predictive model is accurate but takes too long to run.
  (b) When you apply a powerful deep learning algorithm to a simple machine learning problem.
  (c) When the model learns specifics of the training data that can't be generalized to a larger data set.
  (d) When you perform hyper-parameter tuning and performance degrades.

(3) What type of machine learning algorithm makes predictions when you have a set of input data and you know the possible responses?
  (a) Supervised learning.
  (b) Supervisory logic.
  (c) Unsupervised learning.
  (d) Deep learning.

(4) __________ is the machine learning algorithms that can be used with labeled data.
  (a) Regression algorithms.
  (b) Clustering algorithms.
  (c) None of the above.

(5) _________ is a disadvantage of decision trees?
  (a) Decision trees are robust to outliers.
  (b) Decision trees are prone to be overfit.
  (c) Both A and B.
  (d) None of the above.

(6) ____ looks at the relationship between predictors and your outcome.
  (a) Regression analysis.
  (b) K-means clustering.
  (c) Big data.
  (d) Unsupervised learning.

(7) What is an example of a commercial application for a machine learning system?
  (a) A data entry system.
  (b) A data warehouse system.
  (c) A massive data repository.
  (d) A product recommendation system.

(8) To predict a quantity value. use ___.
  (a) Regression.

(b) Clustering.
(c) Classification.
(d) Dimensionality reduction.

(9) You work for an insurance company. Which machine learning project would add the most value for the company?
(a) Create an artificial neural network that would host the company directory.
(b) Use machine learning to better predict risk.
(c) Create an algorithm that consolidates all of your Excel spreadsheets into one data lake.
(d) Use machine learning and big data to research salary requirements.

(10) Why is naive Bayes called naive?
(a) It naively assumes that you will have no data.
(b) It does not even try to create accurate predictions
(c) It naively assumes that the predictors are independent from one another.
(d) It naively assumes that all the predictors depend on one another.

Part 2. Mathematics exercises (all exercises are independent)

Exercise 1. (5 points)
We are interested in estimating parameters α, c. We have independent observations $x_1, \dots, x_n$ ($n \in \mathbb{N}^*$), all of density

$$x \in \mathbb{R} \mapsto \text{Pareto}(x|\alpha, c) = \frac{\alpha c^\alpha}{x^{\alpha+1}} \mathbb{1}_{x>c} \,.$$

(1) We suppose the prior on α, c is $p(\alpha, c) = \mathbb{1}_{\alpha, c>0}$. Compute the posterior $p(\alpha, c|x_1, \dots, x_n)$.
(2) Compute $p(\alpha|c, x_1, \dots, x_n)$.

Exercise 2. (5 points) We have vectors $x^{(1)}, \dots, x^{(N)}$ in $\mathbb{R}^D$ ($N > D$). We have $t_1, \dots, t_N$ in $\mathbb{R}$. We are interested in

$$\widehat{w} = \underset{w \in \mathbb{R}^D}{\arg\min} \sum_{i=1}^N (t_i - w^T x^{(i)})^2 \,.$$

We set

$$x^{(i)} = \begin{pmatrix} x_1^{(i)} \\ x_2^{(i)} \\ \vdots \\ x_D^{(i)} \end{pmatrix} , \forall i \,,$$

$$X = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_D^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \dots & x_D^{(N)} \end{bmatrix} .$$

(1) Show that (for all w)

$$\sum_{i=1}^N (t_i - w^T x^{(i)})^2 = \left(\begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{pmatrix} - Xw \right)^T \left(\begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{pmatrix} - Xw \right) .$$

(2) We set

$$\mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{pmatrix} , \mathcal{L}(w) = \sum_{i=1}^N (t_i - w^T x^{(i)})^2 \,.$$

Show that the gradient of $\mathcal{L}$ is

$$\nabla\mathcal{L}(w) = 2(X^T X)w - 2X^T\mathbf{t}\,.$$

(3) We suppose that $X^T X$ is invertible. Find the absolute minimum of $\mathcal{L}$.