EE 212 – Mathematical Foundations for Machine Learning and Data Science

Final Examination (Summer 2020)

INSTRUCTIONS:

- Exam duration:

  Reading-time: 15 minutes
  Writing-time: 3 hours

- We require you to solve the exam in a single time-slot of three hours and fifteen minutes without any external or electronic assistance.

- We encourage you to solve the exam on A4 paper, use new sheet for each question and write sheet number on every sheet.

- Clearly outline all your steps in order to obtain any partial credit.

- The exam is closed book and notes. You are allowed to have two A4 sheet with you with hand-written notes on both sides. Calculators can be used.

- For the sake of completeness, we require you to write the following statement on your first page of submission: *I commit myself to uphold the highest standards of (academic) integrity.*

- If you are ready, please proceed to the next page.

EE 212 – Mathematical Foundations for Machine Learning and Data Science

Final Examination (Summer 2020)

Total marks: 70 points **Writing Time:** 3 hours

Problem 1. (10 pts)

Consider a network of N links, labeled $1, 2, \ldots, n$. A path through the network is a subset of the links. (The order of the links on a path does not matter here.) Each link has a (positive) delay, which is the time it takes to traverse it. We let d denote the n-vector that gives the link delays. The total travel time of a path is the sum of the delays of the links on the path. Our goal is to estimate the link delays (i.e., the vector d), from a large number of (noisy) measurements of the travel times along different paths. This data is given to you as an $N \times n$ matrix P, where

$$P_{ij} = \begin{cases} 1 & \text{link } j \text{ is on path } i \\ 0 & \text{otherwise} \end{cases}$$

and an N-vector t whose entries are the (noisy) travel times along the N paths. You can assume that $N > n$. You will choose your estimate $\hat{d}$ by minimizing the RMS deviation between the measured travel times (t) and the travel times predicted by the sum of the link delays.

(a) Formulate the problem of determining $\hat{d}$ as least-squares problem.

(b) Provide a matrix expression for $\hat{d}$, that is, solve the least-squares problem and express $\hat{d}$ in terms of P and t.

(c) Explicitly state the assumptions about the data P or t that are required to hold for your expression in part (b).

Problem 2. (**6 pts**)

Three different models are fitted using the same training data set, and tested on the same (separate from the training set) test set (which has the same size as the training set). The RMS prediction errors for each model, on the training and test sets, are reported below. Comment briefly on the results for each model. You might mention whether the model's predictions are good or bad, whether it is likely to generalize to unseen data, or whether it is over-fit.

Model	**Train RMS**	**Test RMS**
A	0.135	2.125
B	0.680	0.675
C	30.03	1.392

Problem 3. (**6 pts**) We consider a following model for the prediction of time-series data $x_1, x_2, \ldots, x_N$

$$\hat{x}_{t+1} = \theta_1 x_t + \theta_2 x_{t-1} + \theta_3 x_t x_{t-1},$$

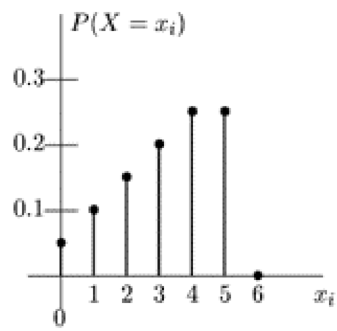
which can be interpreted as 'the value at $t+1$ (future) depends on the values at t (current) and $t-1$ (past)' and this dependence is explained by the above equation and depends on $\theta = (\theta_1, \theta_2, \theta_3)$. Noting that this model can predict x_n for $n \geq 3$, we define the prediction error as

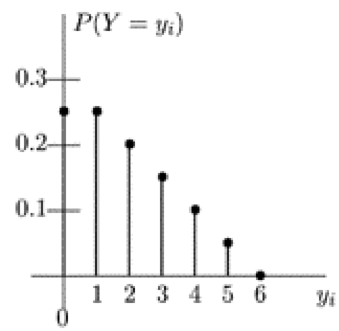
$$\sum_{t=3}^{N-1} \left(\hat{x}_t - x_t\right)^2,$$

that is the sum of square of the error at each time instant. We require you to formulate this error as $\|A\theta - b\|^2$ with $A \in \mathbf{R}^{(N-2)\times 3}$ and $b \in \mathbf{R}^{N-2}$.

Problem 4. **(8 pts)**

Let X be a RV that denotes the number of faulty products in a shipment shipped from Apple, and let Y be the number of faulty products in a shipment shipped from Samsung. The two companies have known probability mass functions which give the probability of a given number of faulty products in a given order. They display their information graphically, as follows:





(a) Suppose you receive a shipment of products from Apple. Find the probability that the number of faulty products is less than or equal to 3.

(b) Suppose you receive a shipment of products from Apple. Somebody tells you that there are at least two faulty products in the shipment. Given this information, find the probability that the number of faulty products is less than or equal to 3.

(c) Suppose a shipment of products arrives, but you are not sure from which company it came; however, you know the probability it came from Apple is 0.35 and the probability it came from Samsung is 0.65. Find the probability that the number of faulty products in the shipment is less than or equal to 3.

Problem 5. (6 pts)

The summer II 2020 term is about to end and the instructor is happy with the performance, claiming that the students have performed above average. A random sample of 30 students' overall scores (marked out of 100) was taken, and the mean was found to be 85.5. The mean score in the previous offerings is 80 with a standard deviation of 12. We are interested in finding out if there is a sufficient evidence to support the instructor's claim.

(a) Write down the null hypothesis H_0 and alternate hypothesis H_a associated with this scenario.

(b) Compute z-statistic for this scenario.

(c) Determine the P-value for this scenario (p-value table is appended with the exam).

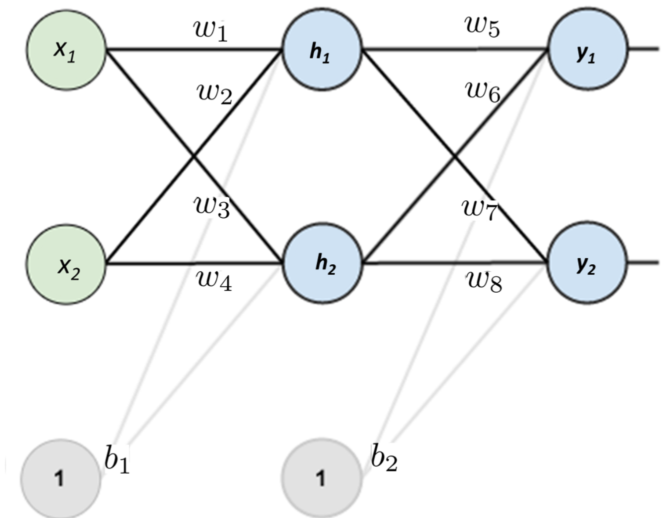
(d) Determine what conclusion can be drawn from the P-value at the following significance level $\alpha = 0.01$.

Problem 6. (4 pts)

An Italian restaurant boasts 320 distinct pasta dishes. Each dish contains exactly one pasta, one meat, and one sauce. If there are 8 pastas and 4 meats available, how many sauces are there to choose from?

Problem 7. (14 pts)

Consider a two-layer neural network with with two inputs, two outputs and two neurons in the hidden layer with interconnections shown in the figure below. We assume that the sigmoid function has been used as the activation function.



(a) Express the output y_1 in terms of inputs x_1 and x_2. You may write a series of equations that relate the output y_1 to the inputs.

(b) We assume that the target outputs are denoted by t_1 and t_2 for the first and second output neuron respectively. Considering the following values for the inputs, weights and biases, determine the outputs y_1 and y_2 and mean-squared-error at the output.

$w_1 = 0.15, \quad w_2 = 0.20, \quad w_3 = 0.25, \quad w_4 = 0.30, \quad b_1 = 0.35$
$w_5 = 0.40, \quad w_6 = 0.45, \quad w_7 = 0.50, \quad w_8 = 0.55, \quad b_2 = 0.60$
$t_1 = 0.01, \quad t_2 = 0.99$
$x_1 = 0.05, \quad x_2 = 0.1$

(c) Derive an expression for the derivative of the error at the output with respect to the weight w_5.

(d) Evaluate the derivative obtained in part(c) using the values and the error computed in part(b).

(e) Assuming the learning rate of 0.01, provide the expression for updating the weight w_5 in terms of previous value of the weight and derivative of the error with respect to the weight. Also compute the value of the updated weight w_5.

Problem 8. (6 pts)

Consider a function $f : \mathbf{R}^n \to \mathbf{R}$ given by

$$f(x) = x^T P x, \quad x \in \mathbf{R}^n,\ P \in \mathbf{R}^{n\times n}$$

(a) Determine the gradient of the function f.

(b) Determine the Hessian matrix of the function f.

(c) What are the conditions on P for a function f to be convex?

Problem 9. (4 pts)

You want to fit a regression model $y = x^T\beta + v$ to the data, using least squares without regularization and you are using QR factorization to compute the model coefficients β and v. Your class-fellow suggests to improve the model by adding a new feature $x = (x, \bar{x})$, where $\bar{x}$ is the average value of x. Note that the added feature is the average of the original features. Choose one of the following, and provide brief justification to support your answer.

(a) This is a good suggestion, and it should be accepted

(b) This is a bad suggestion, and will lead to trouble

Problem 10. (**6 pts**)

Your friend has developed a classifier using a single perceptron for n-vector input and (obviously) a scalar output. Once the classifier is developed using the training data, we have $w \in \mathbf{R}^n$ and $b \in \mathbf{R}$ as weights and bias of the perceptron respectively.

(a) Provide a mathematical expression that relates the input x and output y of the classifier.

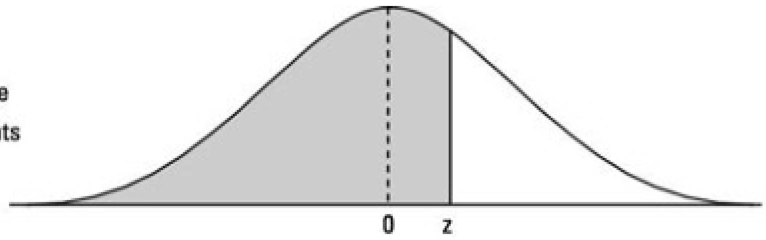
(b) The classifier, once designed, is evaluated on a given test data set and the false positive rate (FPR), defined as the the fraction of the test data points with $y = 0$ for which $\hat{y} = 1$, is computed. Which of the following statements are true (or false)? Provide brief justification to support your answer.

(i) Replacing b with zero will reduce, or not increase, the FPR.

(ii) Replacing w with zero will reduce, or not increase, the FPR.

— End of Exam —

Number in the table represents $P(Z \leq z)$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999