Name:

USC ID:

Notes:

- Write your name and ID number in the solution you submit.
- No books, cell phones or other notes are permitted. Only one letter size cheat sheet (back and front), a tablet for writing down your solutions, and and a calculator are allowed.
- Problems are not sorted in terms of difficulty. Please avoid guess work and long and irrelevant answers.
- Show all your work and your final answer. Simplify your answer as much as you can.
- Open your exam only when you are instructed to do so.
- The exam has 5 questions, 9 pages, and 13 points extra credit.

Problem	Score	Earned
1	22	
2	25	
3	22	
4	22	
5	22	
Total	113	

1. Assume that we built a linear regression model with $p = 5$ predictors. Determine the minimum number of observations n for which $\hat{\beta}_5 = 13.05$ is statistically significant when $\mathrm{SE}(\hat{\beta}_5) = 5.00$. Consider $\alpha = 0.05$.

2. Choose either T (True) or F (False) (no need to explain why):

   (a) When the assumption of conditional independence of features holds, the Naïve Bayes' classifier provides the best accuracy among all possible classifiers. T F

   (b) The F1 score is not an appropriate measure for evaluating binary classifiers when data are not imbalanced. T F

   (c) Leave-One-Out Cross Validation has less bias in estimating the error of a classifier for a large data set than 5 fold cross validation. T F

   (d) When classifying imbalanced data into two classes, we can decrease the threshold on class conditional probability $\Pr(Y = k|X_1 = x_1, \ldots, X_p = x_p]$ to increase the true positive rate at the expense of increasing the false negative rate. T F

   (e) Logistic regression assumes that the conditional odds of the outcome Y given the features, $\mathbb{O}[Y = k|X_1 = x_1, \ldots, X_p = x_p]$, is a logistic function of the features. T F

3. Assume that in a binary classification problem with one feature X, the distribution of X in class $k = 1$ is

$$f_1(x) = \frac{x}{\sigma_1^2} \exp\left(\frac{-x^2}{2\sigma_1^2}\right), x \geq 0$$

and the distribution of X in class $k = 2$ is

$$f_2(x) = \frac{1}{x\sqrt{2\pi}\sigma_2} \exp\left(\frac{-(\ln x - \mu_2)^2}{2\sigma_2^2}\right), x \geq 0$$

(a) Are there any conditions under which the discriminant function is a linear function of x?

(b) If $\sigma_1 = \sigma_2 = 1$, $\mu_2 = 10$, and $\pi_1 = \pi_2 = 0.5$, in what class will $x = 10$ be classified?

4. Consider multinomial regression for multiclass classification with three features $\mathbf{X} = (X_1, X_2, X_3)$, formulated by

$$p_k(\mathbf{X}) = \frac{e^{\beta_{0k}+\beta_{1k}X_1+\beta_{2k}X_2+\beta_{3k}X_3}}{1 + e^{\beta_{01}+\beta_{11}X_1+\beta_{21}X_2+\beta_{31}X_3} + e^{\beta_{02}+\beta_{12}X_1+\beta_{22}X_2+\beta_{32}X_3}}, \; k \in \{1, 2\}$$

where the classes are determined by $k \in \{1, 2, 3\}$

Assume that using a data set of 210 observations from three classes, we obtained the following results:

Coefficient	Value
β_{01}	1
β_{11}	-2
β_{21}	-1
β_{31}	1
β_{02}	0
β_{12}	0
β_{22}	1
β_{32}	1

Assume that the coefficients are all statistically significant.

(a) In what class will the classifier classify $\mathbf{X}^* = (1, 0, -1)$?

(b) Explain why despite having three classes, we formulated multinomial regression using ONLY TWO sets of parameters, $(\beta_{01}, \beta_{11}, \beta_{21}, \beta_{31})$ and $(\beta_{02}, \beta_{12}, \beta_{22}, \beta_{32})$, specific to classes 1 and 2, respectively?

5. In a weird simulated world, we have three types of creatures. Each creature has between 1 to 100 legs, 1 to 100 teeth, and 1 to 100 noses. The fraction of creatures type-1, type-2, and type-3 are respectively $l/(l+t+n)$, $t/(l+t+n)$, and $n/(l+t+n)$, where l, n, t are respectively the number of legs, teeth, and noses of a creature.

   (a) If a creature has 10 teeth, 25 legs, and 30 noses, what is your best guess about the type of the creature?

   (b) What type of supervised learning problem are you solving in this question? Explain.

Scratch paper

Name:

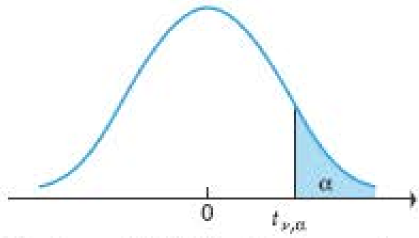
USC ID:

Scratch paper

Name:

USC ID:

Upper Critical Values of Student's t Distribution with ν Degrees of Freedom



For selected probabilities, α, the table shows the values $t_{\nu,\alpha}$ such that $P(t_\nu > t_{\nu,\alpha}) = \alpha$, where t_ν is a Student's t random variable with ν degress of freedom. For example, the probability is .10 that a Student's t random variable with 10 degrees of freedom exceeds 1.372.

	Probability of Exceeding the Critical Value					
ν	0.10	0.05	0.025	0.01	0.005	0.001
1	3.078	6.314	12.706	31.821	63.657	318.313
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.782
8	1.397	1.860	2.306	2.896	3.355	4.499
9	1.383	1.833	2.262	2.821	3.250	4.296
10	1.372	1.812	2.228	2.764	3.169	4.143
11	1.363	1.796	2.201	2.718	3.106	4.024
12	1.356	1.782	2.179	2.681	3.055	3.929
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
60	1.296	1.671	2.000	2.390	2.660	3.232
100	1.290	1.660	1.984	2.364	2.626	3.174
∞	1.282	1.645	1.960	2.326	2.576	3.090
ν	0.10	0.05	0.025	0.01	0.005	0.001