Name:

USC ID:

Notes:

- Write your name and ID number in the spaces above.
- No books, cell phones or other notes are permitted. Only one letter size cheat sheet (back and front) 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 20 points extra credit. However, your grade cannot exceed 100/100.
- In online exams, legible copies SCANNED via phone applications must be submitted, not pictures of answer sheets.
- Make sure you submit ALL pages of your answers. Answers submitted after the exam is adjourned WILL NOT BE ACCEPTED.

Problem	Score	Earned
1	25	
2	30	
3	25	
4	20	
5	20	
Total	120	

1. A regression analysis task relating test scores (Y) to leisure hours (X) produced the following fitted model: $\hat{y} = 25 - 0.5x$.

   (a) What is the fitted value of the response variable corresponding to $x = 7$?

   (b) What is the residual corresponding to the data point with $x = 3$ and $y = 30$?

   (c) If x increases 3 units, how does $\hat{y}$ change?

   (d) An additional test score is to be obtained for a new observation at $x = 6$. Would the test score for the new observation necessarily be 22? Explain.

   (e) The residual sums of squares (RSS) for this model was found to be 7. If there were $n = 16$ observations, provide an estimate for $\sigma^2 = \text{Var}(\epsilon)$.

   (f) If the standard error for $\hat{\beta}_1$ was calculated to be 0.1. Is the predictor statistically significant?

   (g) Rewrite the regression equation in terms of z where z is training time measured in seconds plus one second. Is this new predictor statistically significant. You must argue using statistics, not subjectively.

2. The following dataset was collected to classify people who evade taxes:

Tax ID	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	122 K	No
2	No	Married	77 K	No
3	No	Married	106 K	No
4	No	Single	88 K	Yes
5	Yes	Divorced	210 K	No
6	No	Single	72 K	No
7	Yes	Married	117 K	No
8	No	Married	60 K	No
9	No	Divorced	90 K	Yes
10	No	Single	85 K	Yes

Considering relevant features in the table (only one feature is not relevant), assume that the features are *conditionally independent*. Determine to what class the Naïve Bayes' classifier assigns the test point (Yes, Married, 70K). Assume Gaussianity for continuous features and use Laplace correction for discrete features.

3. Consider the following dataset:

Index	X_1	X_2	Y
1	0	0	1
2	1	0	2
3	0	1	3
4	1	1	2

We wish to fit a multinomial regression model to this dataset.

(a) Write down the likelihood function for this dataset. Use one parameter vector $(\beta_{0k}, \beta_{1k}, \ldots, \beta_{pk}), k = 1, 2, 3$ for each class.

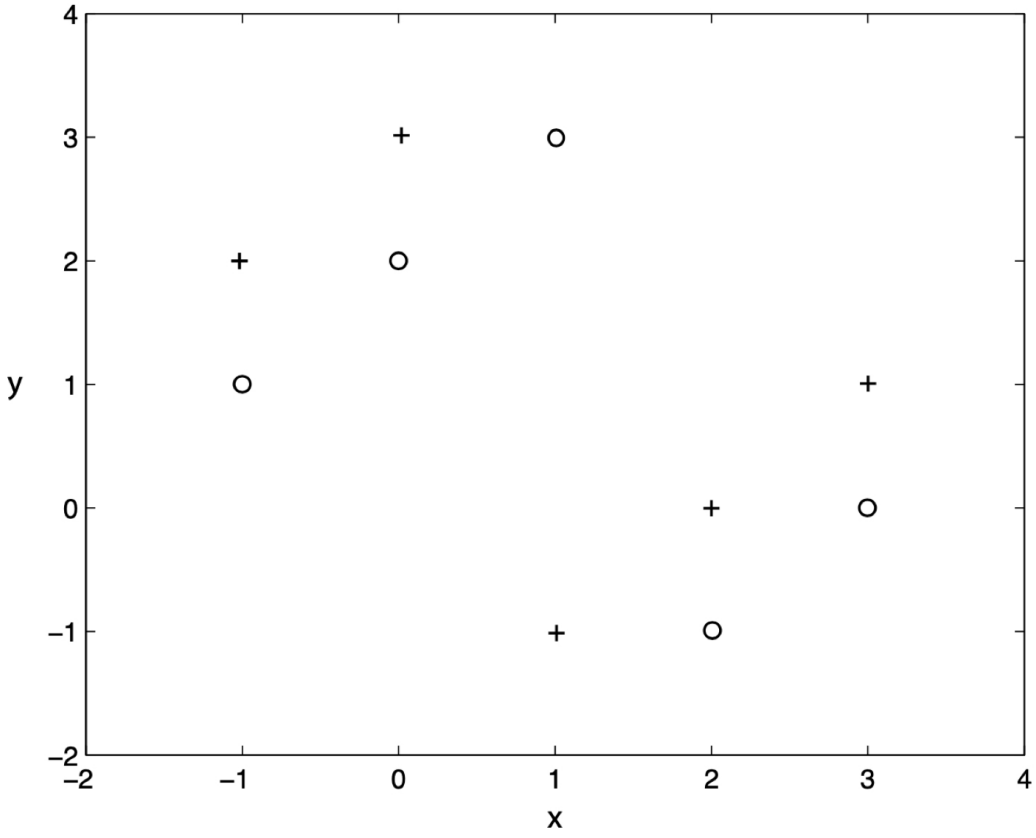
(b) Does this problem have a unique set of β's as its solution?

4. Consider the training set $\{2, 3, 8\}$. There are 10 *distinct* bootstrap samples with the same size that can be drawn from this training set. (For example, $\{2, 3, 3\}$ is not distinct from $\{3, 3, 2\}$). Using those bootstrap samples, build an 80% bootstrap confidence interval for the range, which is defined as the maximum of data minus their minimum.

5. Consider K-NN using Euclidean distance on the following data set (each point belongs to one of two classes: + and ∘).

   (a) What is the leave one out cross validation error when using 1-NN?

   (b) Which of the following values of k leads to the minimum leave-one-out cross validation error: 3, 5 or 9? What is the error for that k? (If there is a tie, please elaborate)



Scratch paper

Name:

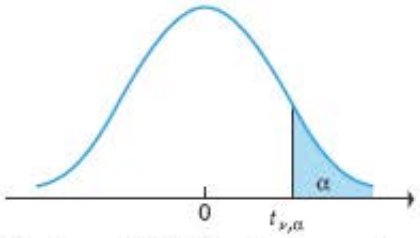
USC ID:

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