# CSE 250B Quiz 2

Tuesday January 21, 2014

*Instructions.* Do this quiz in partnership with exactly one other student. Write both your names at the top of this page. Discuss the answer to the question with each other, and then write your joint answer below the question. Use the back of the page if necessary. It is fine if you overhear what other students say, because you still need to decide if they are right or wrong. You have seven minutes. The maximum score is three points.

*Question.* In a general log-linear model, the denominator is a sum over all possible outcomes of the label $Y$. For standard logistic regression, there are two possible outcomes, $Y = 0$ and $Y = 1$, and the general model is

$$p(Y = 1|x; w) = \frac{\exp \sum_{j=0}^{d} w_{1j} x_j}{(\exp \sum_{j=0}^{d} w_{1j} x_j) + (\exp \sum_{j=0}^{d} w_{0j} x_j)}.$$

The number of parameters in this model is $2(d + 1)$. However, the logistic regression model as seen before is

$$p(Y = 1|x; \beta) = \frac{1}{1 + \exp - \sum_{j=0}^{d} \beta_j x_j}$$

which has only $d + 1$ parameters. Show that the two models are in fact the same.

*Answer.* The first model can be rewritten as

$$\begin{aligned} p(Y = 1|x; w) &= \frac{1}{1 + \frac{\exp \sum_{j=0}^{d} w_{0j} x_j}{\exp \sum_{j=0}^{d} w_{1j} x_j}} \\ &= \frac{1}{1 + \exp \sum_{j=0}^{d} (w_{0j} - w_{1j}) x_j} \end{aligned}$$

which is the same as the second model after the definition $\beta_j = -(w_{0j} - w_{1j})$.

*Additional note.* The first model has redundant parameters: there are many different values for the $w_{0j}$ and $w_{1j}$ parameters that always yield identical predictions. Without regularization, the outcome of training the first model is undefined. The technical term for this situation is that the first model is not identifiable. However, the first model is extended easily to more than two values for the label $y$, and it is identifiable if one uses quadratic regularization.