1. We considered the option of using cross-validation to get multiple estimates of generalization performance of an algorithm in the case when performance cannot be estimated by treating each example independently (e.g., the F1-measure). However, the methods we use to determine the significance of differences in performance make an assumption that estimates are independent. Do estimates derived from cross-validation violate this independent assumption? If so, how? If not, why not?

2. Design a learning algorithm that maximizes area under the ROC curve in a binary classification task.

3. Show that logistic loss (Equations 6.5 on p87 of CIML) is convex for a fixed value of $y \in \pm 1$ and as a function of $\hat{y}$. It’s easiest (shortest, least cumbersome) to do in terms of derivatives, but you could also do it directly from the definition of convexity in terms of chords if you prefer.