

HW04: Generative/Discriminative

Hand in at: <http://www.cs.utah.edu/~hal/handin.pl?course=cs5350>. Remember that only PDF submissions are accepted. We encourage using L^AT_EX to produce your writeups. See `hw01.tex` for an example of how to do so. You can make a `.pdf` out of the `.tex` by running “`pdflatex hw04.tex`”.

1 PRML Exercises

1. *4.9
2. *4.10
3. 4.11 (6350 only)

2 Additional Exercises

1. * Suppose that we train a linear classifier (eg., logistic regression or hinge regression) with an ℓ_2 regularizer on the weights. That is, we minimize $\sum_n \log(1 + \exp[-y_n \mathbf{w}^\top \mathbf{x}_n]) + \frac{\lambda}{2} \|\mathbf{w}\|^2$. (I've left off the bias because it's irrelevant for this question.) I end up with some optimal weight vector which I'll call $\hat{\mathbf{w}}$. Now, suppose that I change all of my inputs by *duplicating* the first feature. That is, if I used to have 100 features, I now have 101 features where feature 1 and feature 2 are *identical*. Now, I relearn weights and compute an optimal set $\hat{\mathbf{w}}'$. How do $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}'$ relate? In particular, how does \hat{w}_1 relate to \hat{w}'_1 and \hat{w}'_2 ? What does this tell us about using feature magnitudes as a measure of feature relevance?
2. (6350 only) What happens to the above question if I use an ℓ_1 regularizer instead of an ℓ_2 regularizer?