

## HW04: Feature and model selection

Hand in via moodle at: <https://moodle.umass.edu/course/view.php?id=20836>. Remember that only PDF submissions are accepted. We encourage using L<sup>A</sup>T<sub>E</sub>X to produce your writeups. See `hw00.tex` for an example of how to do so. You can make a `.pdf` out of the `.tex` by running “`pdflatex hw00.tex`”. You’ll need `mydefs.sty` and `notes.sty` which can be downloaded from the course page.

1. For each of { centering, variance scaling } and each of { decision trees, kNN, perceptron }, state whether the given preprocessing will affect the classifier or not.
2. Assume you have  $D$  features each generated from a zero mean and unit variance Gaussian distribution. In other words, let  $\mathbf{u}$  and  $\mathbf{v}$  are two such vectors such that  $u_i \sim N(0, 1)$ ,  $v_i \sim N(0, 1)$ , show that the quantity

$$\left[ \frac{\|\mathbf{u} - \mathbf{v}\|^2}{D} \right] \rightarrow 2, \text{ as } D \rightarrow \infty.$$

Here  $\rightarrow 2$  implies that the value is tightly concentrated around 2. (Hint: use law of large numbers<sup>1</sup>)

3. For a perceptron you might obtain a confidence value on the prediction by looking at how far the point is from the boundary. How might you obtain confidence values from a decision tree and kNN classifier?
4. Give one reason why 10-fold cross-validation might be preferable over leave-one-out validation?

---

<sup>1</sup>[http://en.wikipedia.org/wiki/Law\\_of\\_large\\_numbers](http://en.wikipedia.org/wiki/Law_of_large_numbers)