6.892 Machine learning seminar

Prof. Tommi S. Jaakkola
tommi@ai.mit.edu

February 8, 1999

This assignment is due **February 15**. You can either email the assignment to me as plain text (use subject "HOMEWORK1") or give it to me in class.

1. Explain in your own words the key ideas of a proof of the following Rao-Blackwell theorem. Be as brief as you like.

   Theorem 1 *Let Y be a scalar random variable and X any vector valued random variable. Assume $E\{Y\} = \mu$ and let $\phi(X) = E\{Y|X\}$. Then*

   (i) $E\{\phi(X)\} = \mu$

   (ii) $Var\{\phi(X)\} \leq Var\{Y\}$

   Here $E\{\cdot\}$ and $Var\{\cdot\}$ are the expected value and the variance, respectively.

2. How would you extend this to the case where Y is a vector valued random variable? Reformulate the theorem and again explain the key ideas of the proof, if any different from above[1].

3. The case of general convex loss functions: Let $\mathrm{Loss}(Y)$ be a real valued loss associated with vector Y. Assume that the loss function is a convex function of Y. Extend the theorem and the proof to the case where "(ii)" is replaced with the expected loss:

   $$(ii) \quad E\{\mathrm{Loss}(\phi(X))\} \leq E\{\mathrm{Loss}(Y)\}$$

   Again, simply give the key ideas.

4. **Research level:** (this is due whenever you are done with it...)

   The wavelet transformation provides multi-resolution representations for time series signals, images, etc.. Derive an analogous method for text documents by appropriately representating elementary text units such as phrases, words, or letter combinations to be able to apply the standard multiresolution transformations such as

[1] Hint: use $tr\Sigma$ where Σ is the covariance matrix of either Y or $\phi(X)$; "tr" stands for "trace".

wavelets. Hint: try mapping (estimating the mapping of) the elementary text units into fairly low dimensional vector space so that the distance between any two points in this representation reflects “semantic distances”. Application area: information retrieval.