This homework is to be done in groups of 3. Each group members should completely understand the group’s solutions and must acknowledge all sources of inspiration, techniques, and/or helpful ideas (web, people, books, etc.) other than the instructor, TA, and class text. Each group should submit a single set of solutions containing the names and e-mail addresses of all group members. Although there are no points for “neatness”, the TA may deduct points for illegible or poorly organized solutions.

1. (3 pts) Problem 3.3 in Bishop.

2. (2 pts) This problem considers the interaction of sample size with the prior (as implemented by regularization). Recall the total error function with regularization, (3.27) in the text:

\[
\frac{1}{2} \sum_{n=1}^{N} (t_n - w^T \phi(x_n))^2 + \frac{\lambda}{2} w^T w.
\]

Consider learning on a set of training points with just one feature that is always 1 and a label that is always 10 (Thus \(\phi(x_n)\) is the scalar 1 and \(t_n = 10\) for all \(N\) training points). First calculate the minimizing \(w\) as a function of \(N\) and \(\lambda\) in this case (do not use a \(w_0\) weight for a constant term, so that \(w\) is simply a scalar).

For what value of \(\lambda\) (as a function of \(N\)) is the minimizing \(w = 5\)?

For fixed \(\lambda\), what happens as \(N\) goes to infinity?

Recommended Problems (not to be turned in):

1. Problem 3.4 in Bishop.