Stat 406 Spring 2007: Homework 5

Out Fri 2 March, back Fri 9 March

1 Ridge regression derivation

Prove that the MAP estimate for linear regression with a Gaussian prior on the weights

$$\log p(\mathbf{w}, \mathcal{D}) = \log \mathcal{N}(\mathbf{w}|0, {\lambda_w}^{-1} I_d)\mathcal{N}(\mathbf{y}|X\mathbf{w}, {\lambda_y}^{-1} I_n) \tag{1}$$

is given by

$$\hat{\mathbf{w}}_{ridge} = \arg\max_{\mathbf{w}} \log p(\mathbf{w}, \mathcal{D}) = (X^T X + \lambda I) X^T \mathbf{y} \tag{2}$$

where $\lambda = \frac{\lambda_w}{\lambda_y}$.

2 Ridge regression and SVD

By performing an SVD decomposition of the design matrix, $X = UDV^T$, prove that

$$\hat{\mathbf{w}}_{ls} = V D^{-1} U^T \mathbf{y} \tag{3}$$

$$\hat{\mathbf{w}}_{ridge} = V(D^2 + \lambda I)^{-1} D U^T \mathbf{y} \tag{4}$$

3 Ridge regression for polynomials (Matlab)

Load the file `sinusoidData`, which contains variables `xtrain10`, `ytrain10`, `xtest` and `ytest`.

1. Fit a polynomial of degree 9 to this data using ridge regression, with $\lambda = 0$, $\lambda = e^{-18}$, and $\lambda = 1$. Plot the resulting fitted functions. You should get something like Figure 1. You can use the provided functions `degexpand`, `standardizeCols` and `ridge` (in the Statistics toolbox).

2. Now run your code using

```
lambdas = [1000 100 10 1 0.1 0.001 1e-4 1e-6 1e-10 1e-12 1e-14];
```

   Plot the root mean squared error on the training and test sets as a function of λ. You should get something like Figure 2.

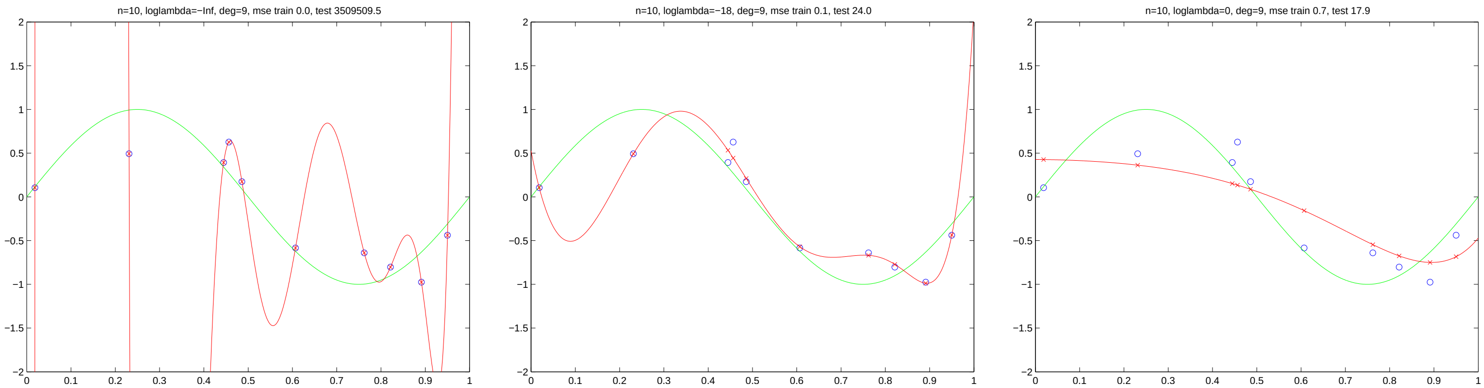

Figure 1: Polynomial regression of order 9 fit with increasing amounts of L2 regularization. Left: $\lambda = 0$ (no regularization). Middle: $\lambda = e^{-18}$. Right: $\lambda = 1$.

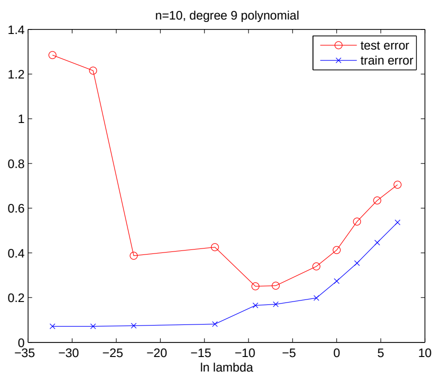

Figure 2: Graph of root mean square error on training ser (lower blue curve) and test set (upper red curve) vs degree of regularization (increase to the right).