School of Electronics and Computer Science
University of Southampton

COMP3206/6229 (2015/16): Machine Learning Lab 5 (of 6) 5%

Issue	Wednesday 11 November 2015
Deadline	Thursday 19 November 2015 (12:00 Noon)

This task is worth 5% of your total assessment. Please work independently. Submit a report of no more than two pages describing your work. Marking Scheme: Three marks for doing correctly what you are asked to do; five marks for showing some insight and/or initiative in completing the tasks.

Construct a radial basis functions (RBF) model to predict house prices using the "Boston Housing Data," used in `Lab 4`. The RBF model is given by

$$g(\boldsymbol{x}) = \sum_{k=1}^{K} \lambda_k \phi(||\boldsymbol{x} - \boldsymbol{c}_k||).$$

We will use a Gaussian RBF $\phi(\alpha) = \exp(-\alpha/\sigma^2)$.

1. Load the data, normalize it as done in `Lab 4` and get random partitions of traing and test sets. Say variable `Xtr`, a matrix of `Ntr` $\times$ `p` is inputs of your training set and `ytr`, the corresponding outputs (targets).

2. Set the widths of the basis functions to a sensibel scale

```
sig = norm(Xtr(ceil(rand*Ntr),:)-Xtr(ceil(rand*Ntr,:));
```

3. Perform $K-$means clustering to find centres for the basis functions. Use $K = Ntr/10$.

```
help kmeans
[C] = kmeans(Xtr, round(Ntr/10))
```

4. Construct the design matrix

```
for i=1:Ntr
    for j=1:K
        A(i,j)=exp(-norm(Xtr(i,:) - C(j,:))/sigma^2);
    end
end
```

5. Solve for the weights

```
lambda = A \ ytr;
```

6. What does the model predict at each of the training data?

```
yh = zeros(Ntr,1);
u  = zeros(Ntr,1);
for n=1:Ntr
   for j=1:K
      u(j)  = exp(-norm(Xtr(n,:) - C(j,:))/sigma^2);
   end
   yh(n) = lambda'*u;
end
plot(y, yh, 'rx', 'LineWidth', 2), grid on
title('RBF Prediction on Training Data', 'FontSize', 16);
xlabel('Target', 'FontSize', 14);
ylabel('Prediction', 'FontSize', 14);
```

7. Adapt the above to calculate what the model predicts at the unseen data (test data) and draw a similar scatter plot. How do the training and test errors compare? Compute the difference between training and test errors at different values of the number of basis functions, K. Briefly comment on any observation you make.

8. Compare your results with the linear regression model of `Lab 4`. Does the use of a nonlinear model improve predictions?

9. Carry out a similar comparison between linear and nonlinear (RBF) models on a different dataset of your choice taken from the `UCI` Machine Learning repository.

Note: When comparing performances of the linear and RBF models, partition the data into training / test sets multiple times (say 20), evaluate the test set prediction errors, and present your results as two boxplots drawn side by side, as in Fig. .

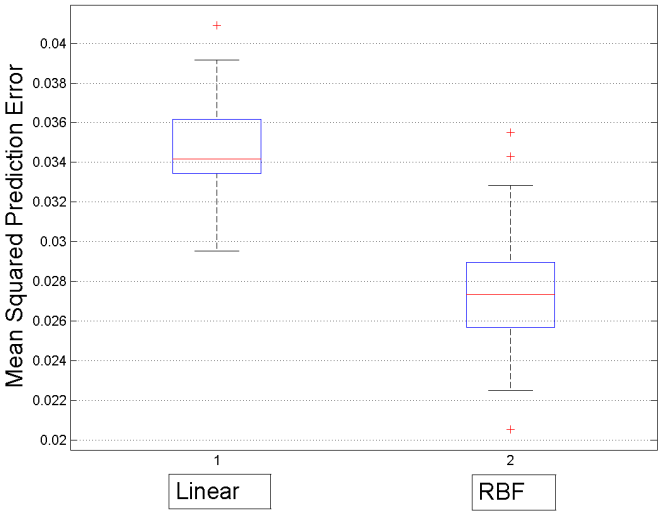

Figure 1: Comparison of linear and radial basis functions (RBF) models on predicting house prices. The mean squared error on test data is obtained from 20 random partitionings.