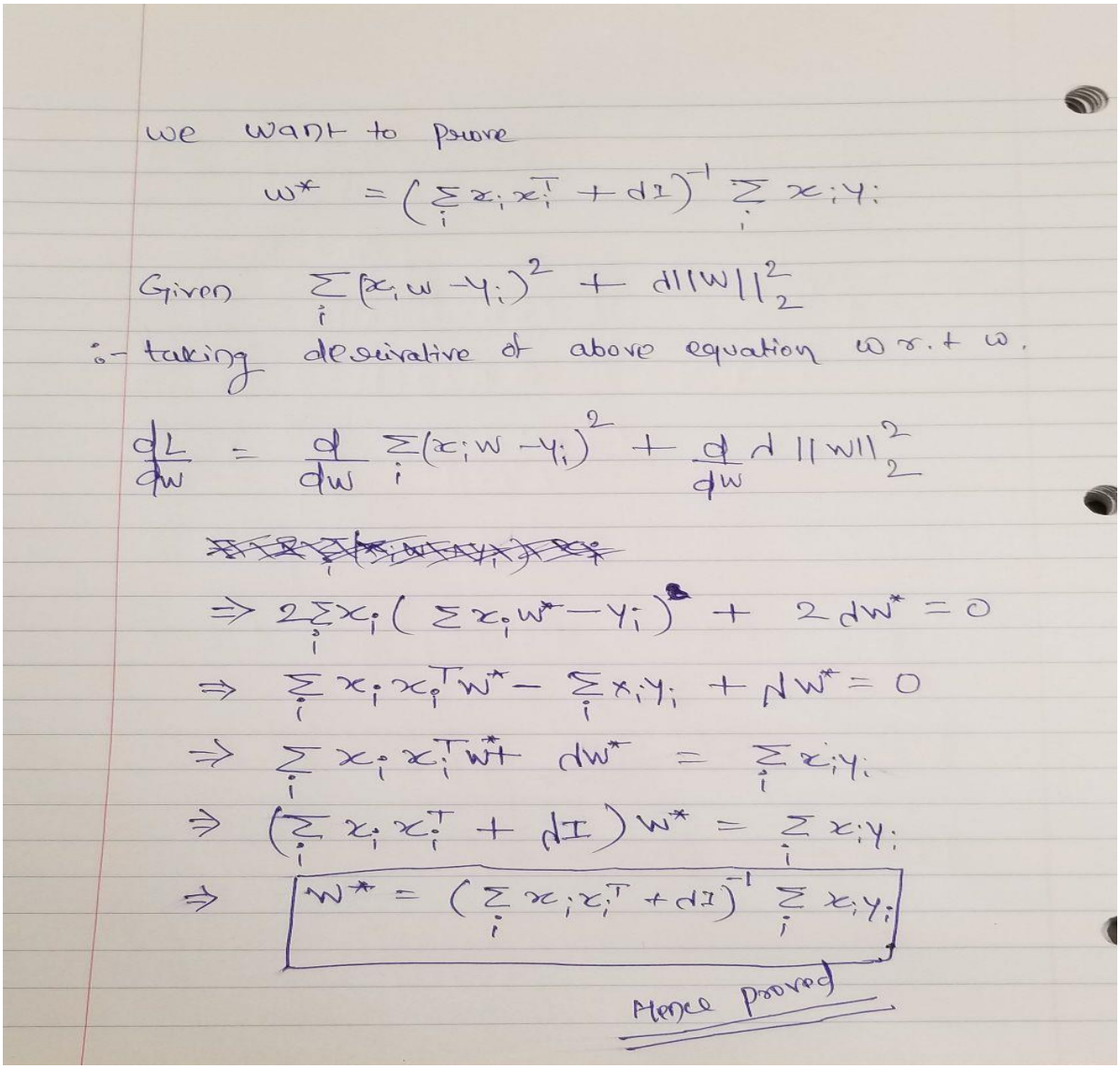
CS589 Machine Learning

Homework 3

Submitted by- Ravi Agrawal Due on- Nov 3rd, 2017

1 A:



we want to prove

$$w^* = \left(\sum_i x_i x_i^T + dI\right)^{-1} \sum_i x_i y_i$$

Given $\sum_i (x_i w - y_i)^2 + d\|w\|_2^2$

∴ taking derivative of above equation w.r.t w.

$$\frac{dL}{dw} = \frac{d}{dw}\sum_i (x_i w - y_i)^2 + \frac{d}{dw} d\|w\|_2^2$$

$$\Rightarrow 2\sum_i x_i \left(\sum x_i w^* - y_i\right) + 2dw^* = 0$$

$$\Rightarrow \sum_i x_i x_i^T w^* - \sum_i x_i y_i + dw^* = 0$$

$$\Rightarrow \sum_i x_i x_i^T w^* + dw^* = \sum_i x_i y_i$$

$$\Rightarrow \left(\sum_i x_i x_i^T + dI\right) w^* = \sum_i x_i y_i$$

$$\Rightarrow \boxed{w^* = \left(\sum_i x_i x_i^T + dI\right)^{-1} \sum_i x_i y_i}$$

Hence proved

1 B:

1B:-

In case of basis expansion x_i will be expanded feature so,

$$\sum_i (\phi(x_i).w - y_i)^2 + \lambda ||w||^2$$

Taking derivate of above equation w.r.t w to find optimal w^*

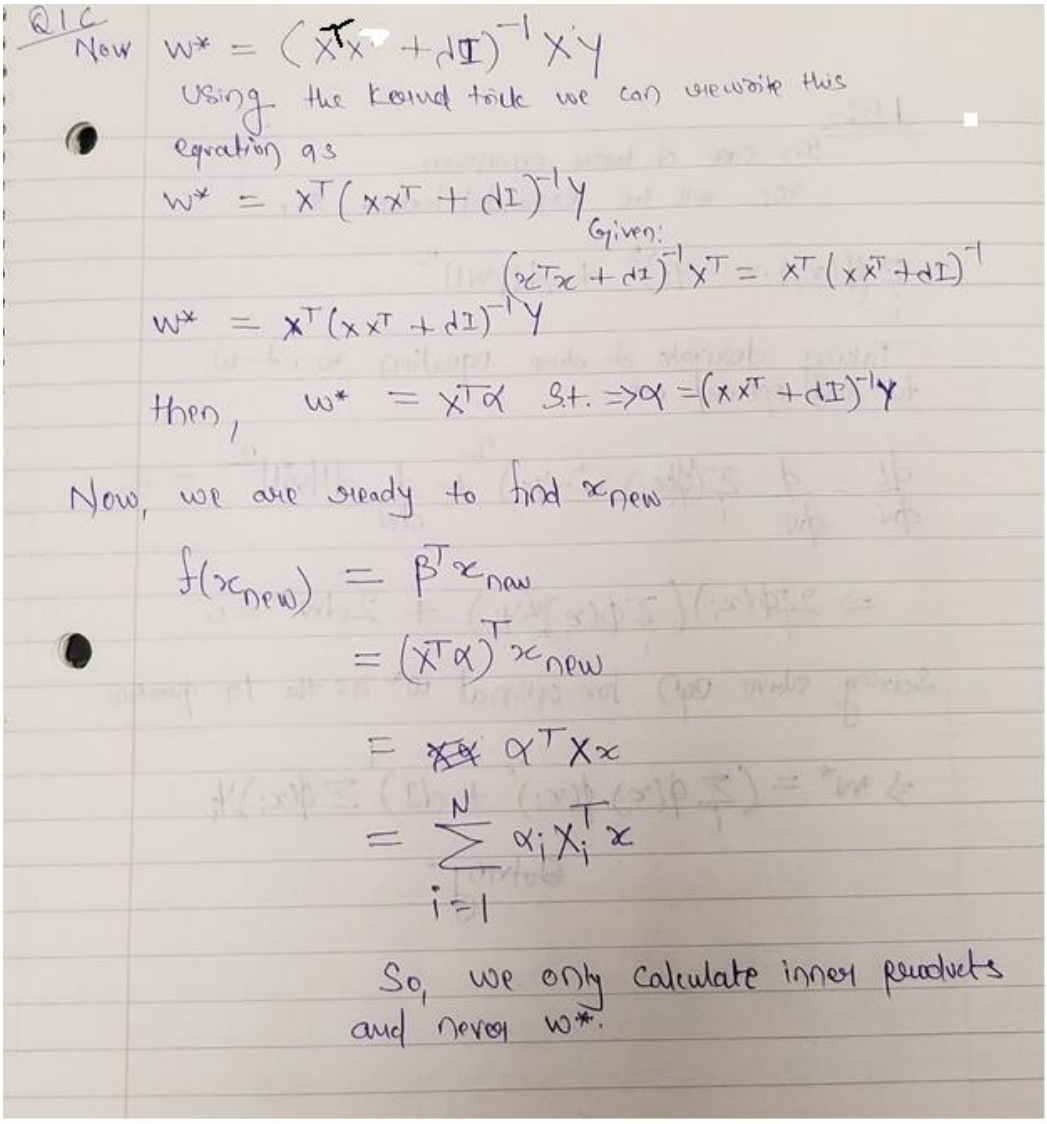
$$\frac{dL}{dw} = \frac{d}{dw} \sum_i (\phi(x_i).w^* - y_i)^2 + \frac{d}{dw} \lambda ||w^*||^2 = 0$$

$$\Rightarrow 2 \sum_i \phi(x_i) \left(\sum_i \phi(x_i) w^* - y_i \right) + 2\lambda w^* = 0$$

Solving above eqn for optimal w^* as the 1a question.

$$\Rightarrow w^* = \left(\sum_i \phi(x_i).\phi(x_i)^T + \lambda I \right)^{-1} \sum \phi(x_i) y_i$$

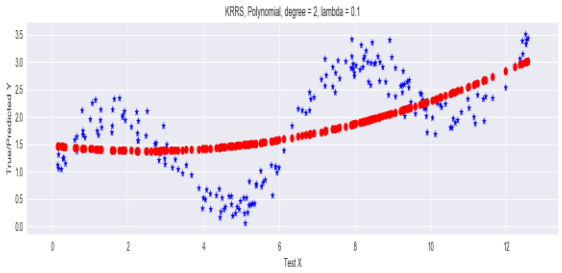
1 C:

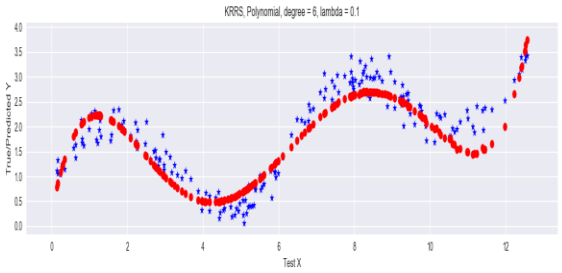


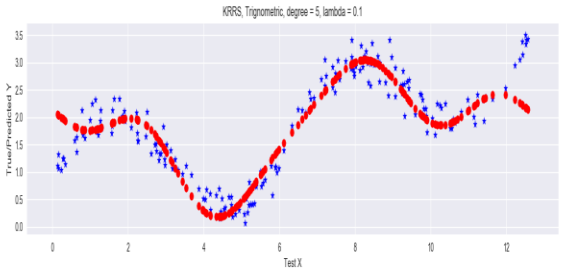
1 d 1:

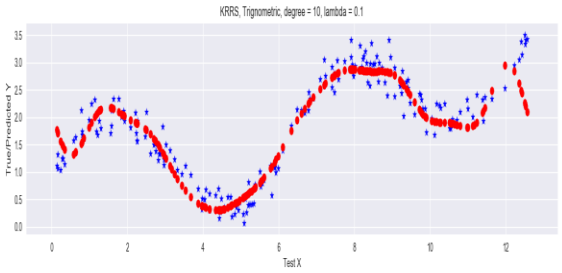
The Plots are displayed in the image below. The most immediate though come after seeing this plots are for the lower order Kernel function or lower degree Basis Expansion functions the model is clearly underfitting and mean squared error is high in such case. Although at the higher order/degree function the Model is generalizing very well and mean squared errors are very low.

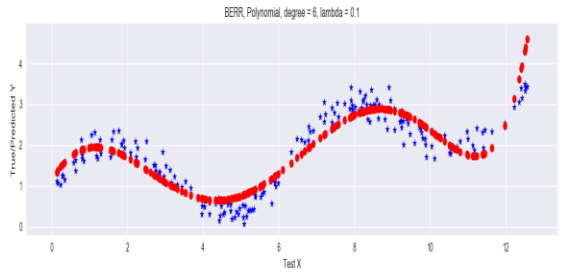
Also, the kernel prediction and the basis expanded feature prediction if one sees the prediction are same.

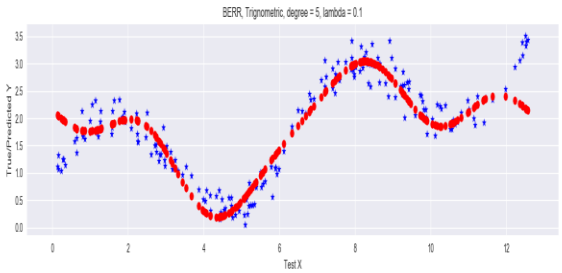


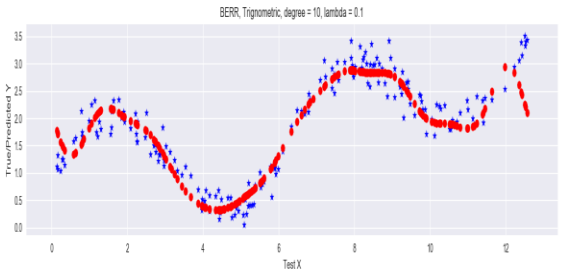












1 d 2: For each of the kernels/basis expansion, the mean squared error is show below : -

	Kernel k(x1, x2)	**Basis Expansion**
Polynomial degree 1	0.582	0.582447550275
Polynomial degree 2	0.536	0.536
Polynomial degree 4	0.441	0.441
Polynomial degree 6	0.118	0.126
Trignometric degree 3	0.165	0.166
Trignometric degree 5	0.130	0.130
Trignometric degree 10	0.097	0.097

1 e: The mean squared error is show in the table below for the three kernels – [RBF, Poly. Degree 3, Linear] with parameters α = [1, 0.0001] and γ = [default, 1, 0.001].

	RBF	Poly. Degree 3	Linear
α = 1, γ = def	14.142	2.484	10.796
α = 1, γ = 1	69.223	2.590	--
α = 1, γ = 0.001	9.367	6.291	--
α =0.0001, γ = def	1.971	2.461	10.492
α = 0.0001, γ = 1	17.470	2.590	--
α = 0.0001, γ = 0.001	2.339	1.987	--

The Best model using the 10 folds cross validation was chosen to train on the full dataset. The RBF kernel model was found to be the best model with parameter α = 0.0001 and γ = def. The best model gives the Mean Squared error of **0.5072** on the Kaggle under the **username- imraviagrawal.** The MSE from the Kaggle leaderboard 0.5072 and model selection generated MSE 1.971 it is clear that, the RBF model is working better in the model selection and the Kaggle as well.

2 a : The out of Sample accuracy for the models is shown in the figure below:

	RBF	Poly. degree 3	Poly. degree 5	Linear
C = 1, γ = 1	0.875	0.944	0.944	0.95923
C = 1, γ = 0.01	0.967	0.959	0.950	--
C = 1, γ = 0.001	0.965	0.914	0.746	--
C = 0.01, γ = 1	0.652	0.944	0.944	0.965
C = 0.01, γ = 0.01	0.933	0.952	0.952	--
C = 0.01, γ = 0.001	0.652	0.652	0.652	--
C = 0.0001, γ = 1	0.652	0.946	0.944	0.946
C = 0.0001, γ = 0.01	0.652	0.701	0.875	--
C = 0.0001, γ = 0.001	0.652	0.652	0.652	--

'C': 1, 'degree': 3, 'gamma': 0.01, 'kernel': 'rbf'

The best model with kernel "RBF" , degree 3 and parameters C equals 1 and gamma equals 0.01 was trained on the full training dataset and prediction was made for the test set. Predicted output was submitted on the **Kaggle** and produced the accuracy score of **0.95689.** The Validation Accuracy for the same model is **0.967** which is almost similar. The SVM on the credit card dataset is working with cross validation.

Extra Credit 3:

I have experimented Kernel ridge regression with the following kernels “rdf” , “poly”, “linear”, and “sigmoid” for the following values of the alpha 0.00001, 0.0001, 0.001, 0.01 and 1, the gamma as None, 0.001, 0.001 and 1 and the degrees 1, 3, 4, and 8. The best model with the rbf kernel with alpha 0.0001 and gamma 0.01 was selected using the gridsearchcv function. The best model was trained on the full dataset and prediction is submitted to the extra credit question leaderboard. The model gives the MSE of 1.179 and the model at the time of the reporting give me position under **top 10** in the leadership board.