CS680 Assignment 3 Kilby Baron - 20773955

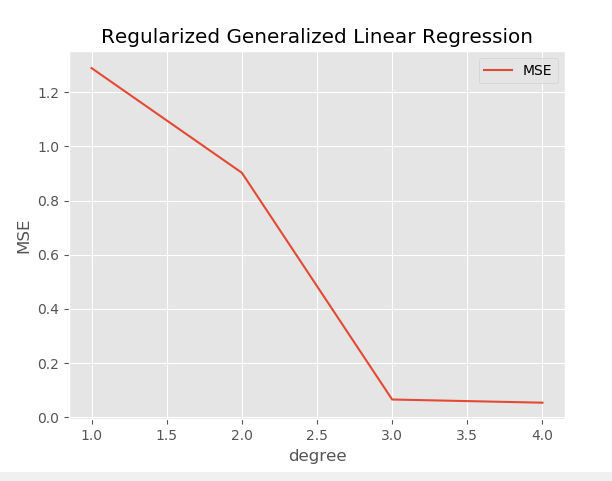
**Q1) Show that k(x,x’) can be expressed as the inner product of an infinite feature space**

*k*(x,x’) = *exp*(-||x-x’||2/2 σ2)

= (*e* −xTx/2σ^2)(*e* xTx’/σ^2)(*e* −(x’)Tx’/2σ^2)

= (*e* −xTx/2σ^2)(*e* −(x’)Tx’/2σ^2)()In the summation of the final factor, the numerator is equivalent to the polynomial kernel of degree n. The polynomial kernel projects a given vector into a higher dimensional space, as demonstrated in question 2c. Since n extends to infinity in this summation, the Gaussian kernel k(x,x’) is the inner product of a feature space with infinite dimensions.

**Q2) Non-linear regression techniques**

1. **Regularized generalized linear regression**

**Best degree:** 4

**Squared error for test set using 4 degrees:** 0.0446196

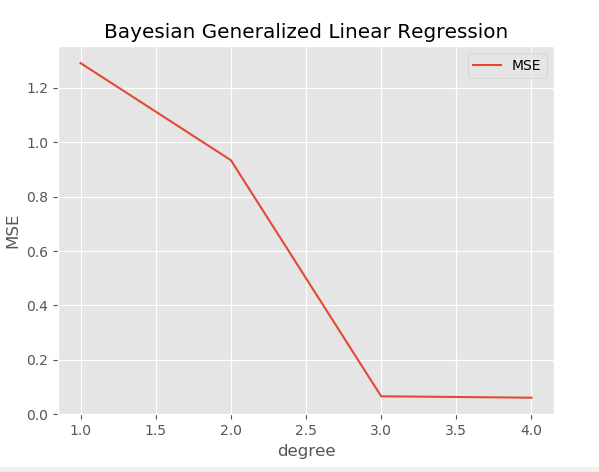
**Running time variance for cross validation:**

degree 1: 0.0052904 s

degree 2: 0.0086279 s

degree 3: 0.0135583 s

degree 4: 0.0188342 s

1. **Bayesian generalized linear regression**

**Best degree:** 4

**Squared error for test set using 4 degrees:** 0.050440

**Running time variance during cross validation:**

degree 1: 0.00521 s

degree 2: 0.00800 s

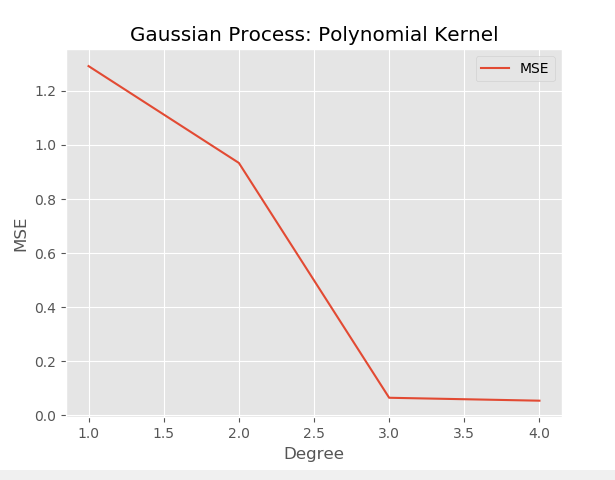
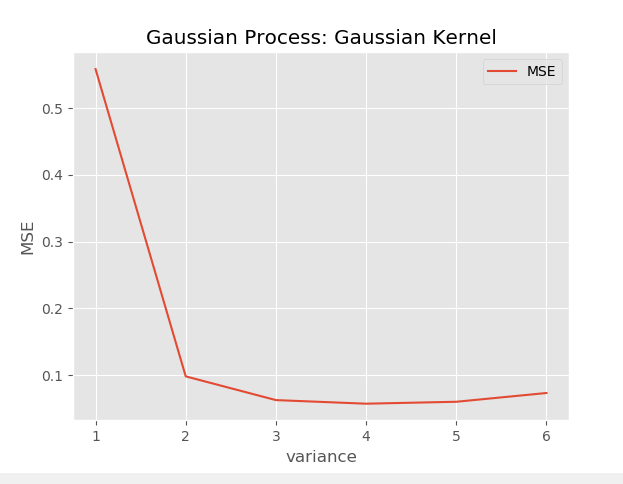
degree 3: 0.01286 s

degree 4: 0.01726 s

**Regularized generalized linear regression vs Bayesian generalized linear regression:**

|  |  |
| --- | --- |
| **Similarities** | **Differences** |
| -Both are non-linear regression models  -Both use non-linear basis functions to add dimensions to data | In Bayesian learning, instead of making predication according to w\*, we compute the weighted average prediction according to Pr(**w**|**X,y**) |

1. **Gaussian process regression**

**Identity Kernel:** Squared error of test set for identity kernel: 3.59062069

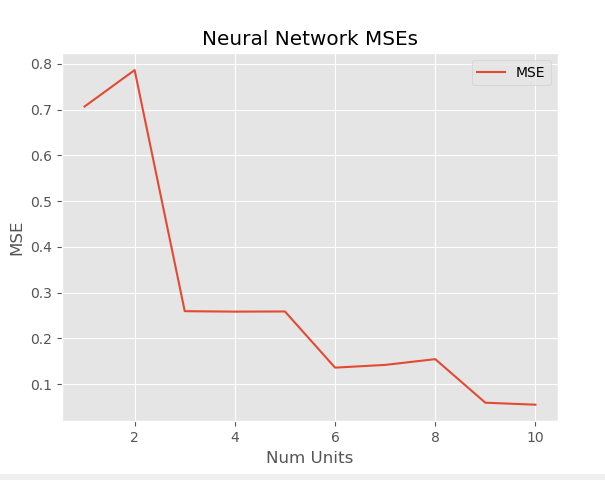
**SE of gaussian kernel using best variance (4):** 0.04302516

**SE of test set for polynomial kernel using best degree (4):** 0.04577037

**Run time variance on test set:**

|  |  |  |
| --- | --- | --- |
| Identity | Gaussian | Polynomial |
| 0.7518 s | 0.9580 s | 0.6426 s |

1. **Neural network**



**Best number of hidden units:** 10

**Squared error of test set using best number of units:** 0.02947

**Run time variance:**

1 unit: 142.51 s

2 units: 161.45 s

3 units: 183.18 s

4 units: 210.26 s

5 units: 244.19 s

6 units: 278.83 s

7 units: 313.85 s

8 units: 348.27 s

9 units: 387.42 s

10 units: 432.11 s