CSC411 Assignment 3

Yue Guo

December 3, 2017

1 20 News Group

1.1 Top 3 Algorithms

1.1.1 Neural Network

1. Hypterparameter

   In the code nn_news, I have tried a single layer neural network vs multi-layered neural network. It turns out that the single neural network is the fastest and also most accurate.

2. Train/test loss

   - Train accuracy
   - Test accuracy

3. My expectations

1.1.2 Random forest

1. Hypterparameter

   In the code nn_news, I have tried a single layer neural network vs multi-layered neural network. It turns out that the single neural network is the fastest and also most accurate.

2. Train/test loss

   - Train accuracy
   - Test accuracy

3. My expectations

1.1.3 SVM - Best Classifier

1. Hypterparameter

   In the code nn_news, I have tried a single layer neural network vs multi-layered neural network. It turns out that the single neural network is the fastest and also most accurate.

2. Train/test loss

   - Train accuracy 0.972511932119
   - Test accuracy 0.691980881572

3. My expectations

1.1.4 Bernoulli Baseline

1. Train/test loss

   - Train accuracy 0.598727240587
   - Test accuracy 0.457912904939

2 SVM

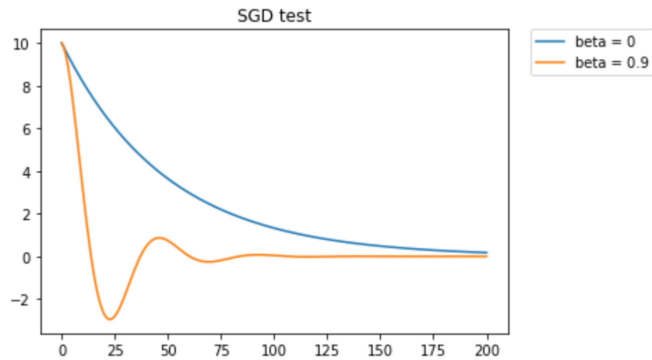

Figure 1: Plot of test SVM

3 Kernels

3.1 Positive semidefinite and quadratic form

Assume K is symmetric, we can decompose K into $U\Lambda U^T$

$$x^T K x = x^T (U\Lambda U^T)x = (x^T U)\Lambda(U^T x)$$

Λ has the eigenvalues λ_i, and if K is positive, and all $\lambda_i > 0$,

$$x^T K x = \Sigma_{i=1}^{d} \lambda_i([x^T U_i])^2 >= 0$$

Then $x^T K x >= 0$ for all x in $\mathbb{R}^d$

3.2 Kernel properties

3.2.1 α

Define mapping $\phi(x) = \sqrt{\alpha}$, $\alpha > 0$, and the kernel $\langle\phi(x), \phi(y)\rangle = \alpha$. The resulting matrix K has item $K_{ij} = \alpha$, the matrix K has equal number of row and columns, and each element is α. Since $\alpha > 0$, and all elements are equal, K is positive semidefinite

3.2.2 $f(x), f(y)$

$K_{ij} = \langle\phi(x), \phi(y)\rangle$,
define $\phi(x) = f(x), \forall f : \mathbb{R}^d \to \mathbb{R}$
define $\phi(y) = f(y), \forall f : \mathbb{R}^d \to \mathbb{R}$
Since f(x) and f(y) produce a scalar, $\langle\phi(x), \phi(y)\rangle = f(x) \cdot f(y)$

3.2.3 k1 and k2

If the gram matrix, K_1 of kernel k1 and gram matrix, K_2 of kernel k2 are positive semidefinite, by scaling them and adding each element, the new gram matrix of $a \cdot k_1(x, y) + b \cdot k_2(x, y)$, call it K, each element of K is positive since a ,b > 0.
K is also symmetric because K_1 and K_2 are symmetric with the same dimension, and element wise addition and linear combination preserve the symmetric property.

3.2.4 $k(x,y) = \frac{k_1(x,y)}{\sqrt{k_1(x,x)}\sqrt{k_1(y,y)}}$

Let ϕ_1 be the mapping defined by k_1
We define a new mapping, ϕ for $k(x,y)$
We let $\phi(x) = \frac{\phi_1(x)}{\|\phi_1(x)\|}$

$$\begin{aligned}
k(x,y) &= \langle \phi(x), \phi(y) \rangle \\
&= \frac{\phi_1(x)}{\|\phi_1(x)\|} \cdot \frac{\phi_1(y)}{\|\phi_1(y)\|} \\
&= \frac{\phi_1(x)}{\sqrt{\phi_1(x) \cdot \phi_1(x)}} \cdot \frac{\phi_1(y)}{\sqrt{\phi_1(y) \cdot \phi_1(y)}} \\
&= \frac{\phi_1(x)}{(\sqrt{\phi_1(x)} \cdot \sqrt{\phi_1(y)})} \cdot \frac{\phi_1(y)}{(\sqrt{\phi_1(x)} \cdot \sqrt{\phi_1(y)})} \\
&= \frac{\phi_1(x)}{\sqrt{\phi_1(x) \cdot \phi_1(y)}} \cdot \frac{\phi_1(x)}{\sqrt{\phi_1(x) \cdot \phi_1(y)}} \\
k(x,y) &= \frac{k_1(x,y)}{\sqrt{k_1(x,x)}\sqrt{k_1(y,y)}}
\end{aligned}$$

Therefore, there is a new mapping $\phi(x)$ that supports $k(x,y)$ and it is a kernel because $\phi(x)$ is the product of two kernel mappings