Machine Learning for Statistics (Spring 2017)
Homework #2

Minxia Ji - mji4@fordham.edu
Discussants:

March 12, 2017

Problem 1

Primal form:

$$minL(\theta, \theta_0, \alpha) = \frac{1}{2}||\theta||^2 - \sum_{i=1}^{m} \alpha_i y_i (x_i \cdot \theta + \theta_0) + \sum_{i=1}^{m} \alpha_i$$

$$s.t. \forall i \quad \alpha_1 \geq 0 \quad wheremisthenumberoftrainingpoints$$

take the derivative of $L(\theta, \theta_0, \alpha)$ with the respect to θ and θ_0

$$\theta = \sum_{i=1}^{m} \alpha_i y_i x_i, \quad \sum_{i=1}^{m} \alpha_i y_i = 0$$

take back to $L(\theta, \theta_0, \alpha)$ and get dual form:

$$max \quad g(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j ({x_i}^T x_j)$$

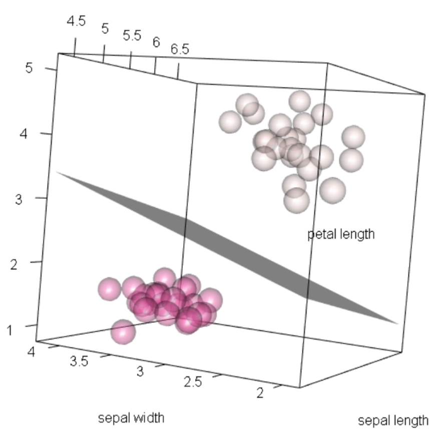
$$s.t. \quad \sum_{i=1}^{m} \alpha_i y_i = 0 and \alpha_i \geq 0$$

Problem 2

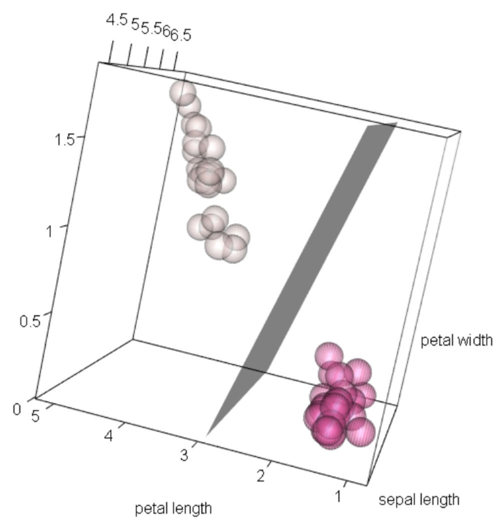
Change the species setosa to the value 1 and the versicolor to the value -1.Sampling the whole data and splitting the data into two data sets, train data and test data. Use the function, I obtained theta0 = 2.2762320, theta = -0.2400101 0.4819709 -0.8642628 -0.6039267 Using theta to predict the accuracy of the model when it is applied to the test data.

Problem 3

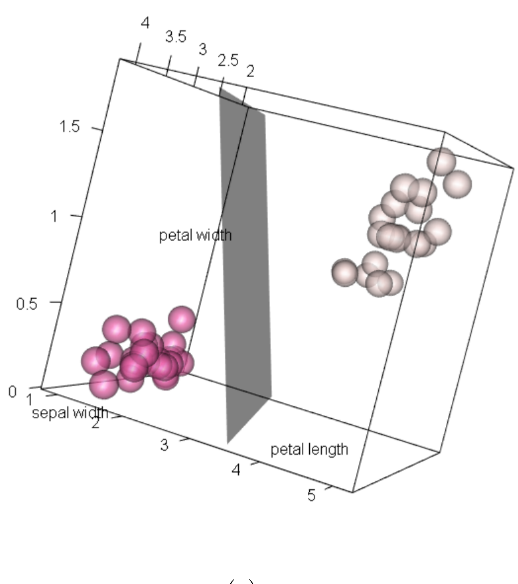
The hot-pink spheres represent setosa and the light-pink spheres represent versicolor.The gray plane represents decision boundary.Each plot was drew by selecting three features from iris data(4 features).Therefore, there are four plots.The features are shown in each plot. The prediction error of my model is 0%.



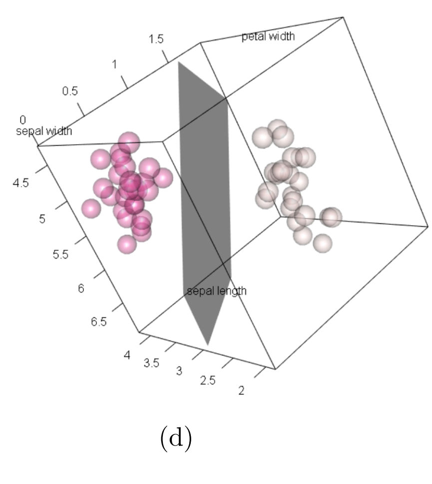
(a)



(b)



(c)



(d)

Problem 4

a).

Letting K denote a Mercer kernel($K_{ij} = \phi(x_i)^T\phi(x_j) = K_{ji}$, symmetric) and $\phi_k(x)$ denote the k-th coordinate of vector $\phi(x)$, we find that for any vector v, we have

$$v^T K v = \sum_i \sum_j v_i K_{ij} v_j$$

$$= \sum_i \sum_j v_i \phi(x_i)^T \phi(x_j) v_j$$

$$= \sum_i \sum_j v_i \sum_k \phi_k(x_i)^T \phi_k(x_j) v_j$$

$$= \sum_k \sum_i \sum_j v_i \phi_k(x_i)^T \phi_k(x_j) v_j$$

$$= \sum_k (\sum_i z_i \phi_k(x_i))^2 \geq 0$$

Since v is arbitrary, this shows that k is positive semi-definite.

b).

Proof $k(x, x') = ak_1(x, x') + bk_2(x, x')$ for $a, b \geq 0$

Letting ϕ_1 and $\langle\rangle_{k_1}$ denote the feature map and inner product for k_1 and ϕ_2 and $\langle\rangle_{k_2}$ denote the feature map and inner product for k_2, by linearity, we have

$$ak_1(x, x') = \langle\sqrt{a}\phi_1(x), \sqrt{a}\phi_1(x')\rangle_{k_1} and bk_2(x, x') = \langle\sqrt{b}\phi_2(x), \sqrt{b}\phi_2(x')\rangle_{k_2}$$

Then,

$$k(x, x') = ak_1(x, x') + bk_2(x, x')$$

$$= \langle\sqrt{a}\phi_1(x), \sqrt{a}\phi_1(x')\rangle_{k_1} + \langle\sqrt{b}\phi_2(x), \sqrt{b}\phi_2(x')\rangle_{k_2}$$

$$= \langle[\sqrt{a}\phi_1(x), \sqrt{b}\phi_2(x)], [\sqrt{a}\phi_1(x'), \sqrt{b}\phi_2(x')]\rangle_{k_{new}}$$

This shows $k(x, x')$ can be expressed as an inner product.

Proof $k(x, x') = k_1(x, x') \times k_2(x, x')$

Let $f_i(x)$ and $g_i(x)$ be the i-th feature value under the feature map ϕ_1 and ϕ_2

Then,

$$k(x, x') = k_1(x, x') \times k_2(x, x')$$

$$= \langle\phi_1(x), \phi_1(x')\rangle\langle\phi_2(x), \phi_2(x')\rangle$$

$$= (\sum_i f_i(x) f_i(x'))(\sum_j g_j(x) g_j(x'))$$

$$= \sum_{i,j} f_i(x) f_i(x') g_i(x) g_i(x')$$

=

$$\sum_{i,j} (f_i(x)g_i(x))(f_i(x')g_i(x'))$$

$$= \langle \phi_3(x), \phi_3(x') \rangle$$

where ϕ_3 has feature $h_{i,j}(x) = f_i(x)g_i(x)$, then $k(x,x') = \phi(x)^T\phi(y)$
Therefore, $k(.,.)$ is a Mercer kernel

Proof $k(x,x') = f(k_1(x,x'))$ where f is a polynomial with positive constants.
Since each polynomial term is a product of kernels with a positive coefficient, the proof follows by applying $k(x,x') = ak_1(x,x') + bk_2(x,x')$ for $a, b \geq 0$ and $k(x,x') = k_1(x,x') \times k_2(x,x')$

Proof $k(x,x') = exp(k_1(x,x'))$
Since using Tailor series expansion $e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + ...$, the proof basically follows from $k(x,x') = f(k_1(x,x'))$.

c).
$K(x,y) = exp(-\frac{1}{2}||x-y||^2) = exp(-\frac{1}{2}x^2 + xy - \frac{1}{2}y^2) = exp(-\frac{1}{2}x^2 - \frac{1}{2}y^2)exp(xy)$
using Tailor series expansion of e^{xy}, we can have,

$$exp(-\frac{1}{2}x^2 - \frac{1}{2}y^2)[1, \frac{1}{1!}xy, \frac{1}{2!}(xy)^2, \frac{1}{3!}(xy)^3, ...]^T$$

which equal to,

$$exp(-\frac{1}{2}x^2)[1, \sqrt{\frac{1}{1!}}x, \sqrt{\frac{1}{2!}}x^2, \sqrt{\frac{1}{3!}}x^3, ...]^T \cdot exp(-\frac{1}{2}y^2)[1, \sqrt{\frac{1}{1!}}y, \sqrt{\frac{1}{2!}}y^2, \sqrt{\frac{1}{3!}}y^3, ...]^T = \phi(x)\phi(y)$$

therefore, we have the function for ϕ as $\phi(x) = e^{-\frac{1}{2}x^2}[1, \sqrt{\frac{1}{1!}}x, \sqrt{\frac{1}{2!}}x^2, \sqrt{\frac{1}{3!}}x^3, ...]^T$

Problem 4(bonus)

I have written functions to calculate Gaussian and polynomial kernel and check the output with R package kernlab. However I don't know how to use solve.qp to solve dual form of SVM and I am still working on that.