CMSE 820: Homework #9

Due on Nov 17, 2019 at 11:59pm

Professor Yuying Xie

Boyao Zhu

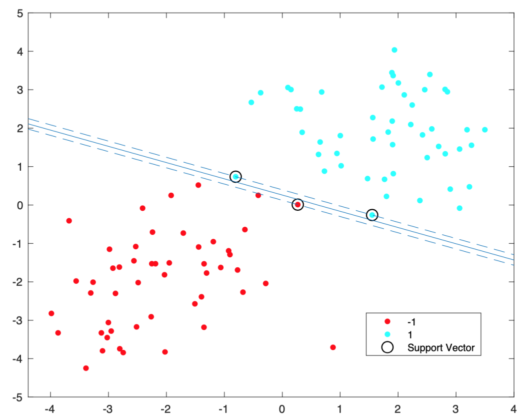

Figure 1: Hard margin SVM on a separable dataset. The colors indicate the original label in the data set

Problem 1

Solution
(a) Figure 1 (b) After makeing the change in the dataset, the dataset becomes non-separable linearly, so hard margin SVM yields no solution. (c) Figure 2

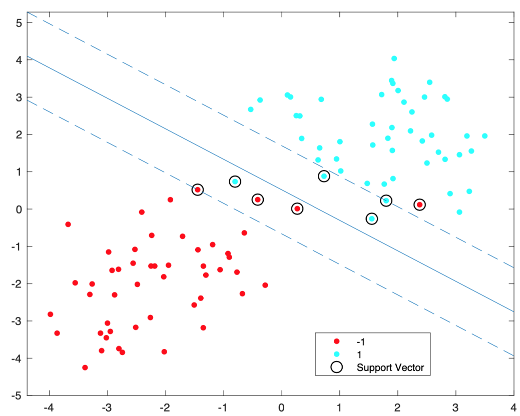

Figure 2: Soft margin SVM on a non-separable dataset. The colors indicate the original label in the dataset.

Problem 2

Solution
Figure 3

Figure 4
Figure 5

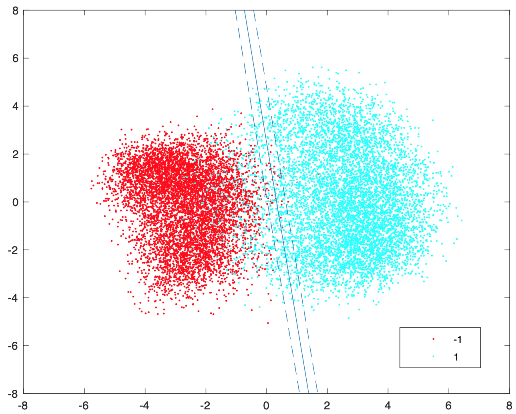

Figure 3: Hard margin SVM on the original dataset. The solid line is the intersection of the SVM hyperplane and the plane spanned by the top 2 PCs.

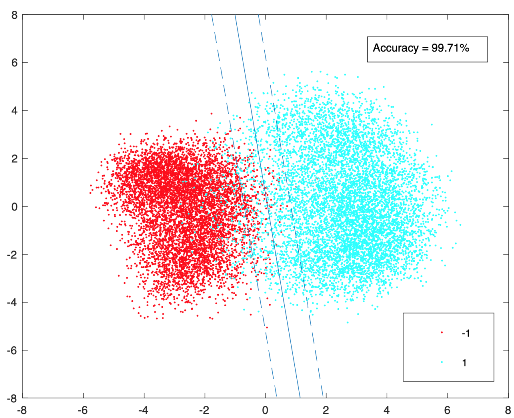

Figure 4: Soft margin SVM on the original dataset. The solid line is the intersection of the SVM hyperplane and the plane spanned by the top 2 PCs.

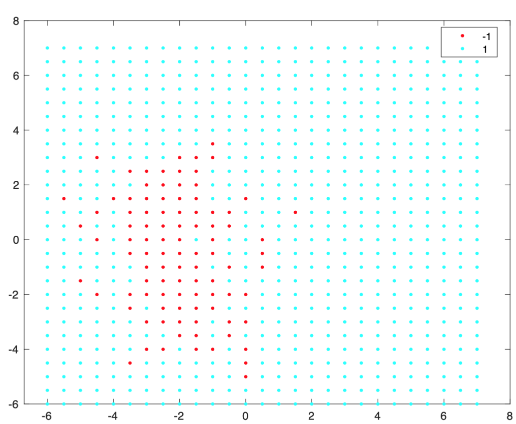

Figure 5: Predictions of the SVM model with a Gaussian kernel on the grid spanned along the top 2 PCs.

Problem 3

Soft Margin SVM
(1) Regularized risk form:

$$\min_{w,b} \sum_i [1 - y_i(\langle w, x_i \rangle + b)]_+ + \frac{\lambda}{2}\|w\|^2$$

(2) Primal form:

$$\min_{w,b} \gamma \sum_i \xi_i + \frac{1}{2}\|w\|^2$$

$$\text{subject to } \xi_i > 0, \quad y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i$$

Show that they are equivalent with $\lambda = \frac{1}{\gamma}$.

Solution
Substituting $\lambda = \frac{1}{\lambda}$, the regularized risk form problem can be rewritten as

$$\min_{w,b} \gamma \sum_i \max\{0, 1 - y_i(\langle w, x_i \rangle + b)\} + \frac{1}{2}\|w\|^2$$

Let $\{w_1, b_1\}$ and $\{w_2, b_2\}$ denote respectively the optimal solutions to the regularized risk form problem and the primal problem respectively.
The Lagrangian for the primal problem reads

$$L(w, b, \xi_i, \alpha_i, \lambda_i) = \gamma \sum_i \xi_i + \frac{1}{2}\|w\|^2 + \sum_i \alpha_i [1 - \xi_i - y_i(\langle w, x_i \rangle + b)] - \sum_i \lambda_i \xi_i$$

The complementary slackness yields

$$\forall i, \lambda_i \xi_i = 0 \text{ and } \alpha_i [1 - \xi_i - y_i(\langle w_2, x_i \rangle + b_2)] = 0$$

So the objective function of the primal problem at the optimal solution $\{w_2, b_2\}$ is given by

$$\gamma \sum_i \max\{0, 1 - y_i(\langle w_2, x_i \rangle + b_2)\} + \frac{1}{2}\|w_2\|^2 \geq \gamma \sum_i \max\{0, 1 - y_i(\langle w_1, x_i \rangle + b_1)\} + \frac{1}{2}\|w_1\|^2$$

The inequality follows from that $\{w_1, b_1\}$ is the optimal solution to the regularized form problem.
On the other hand, given $\{w_1, b_1\}$, we can define $\xi_i' := \max\{0, 1 - y_i(\langle w_1, x_i \rangle + b_1)\}$ such that ξ_i' is feasible for the primal form problem. The optimality of the primal form solution $\{w_2, b_2\}$ gives

$$\gamma \sum_i \max\{0, 1 - y_i(\langle w_2, x_i \rangle + b_2)\} + \frac{1}{2}\|w_2\|^2 \leq \gamma \sum_i \max\{0, 1 - y_i(\langle w_1, x_i \rangle + b_1)\} + \frac{1}{2}\|w_1\|^2$$

We obtain an equality from the previous two inequalities

$$\gamma \sum_i \max\{0, 1 - y_i(\langle w_2, x_i \rangle + b_2)\} + \frac{1}{2}\|w_2\|^2 = \gamma \sum_i \max\{0, 1 - y_i(\langle w_1, x_i \rangle + b_1)\} + \frac{1}{2}\|w_1\|^2$$

Moreover, since the soft margin SVM problem is a strictly convex optimization problem, the solution is unique, i.e.,

$$w_1 = w_2 \text{ and } b_1 = b_2.$$

Thus, the two forms of problems are equivalent.

Problem 4

Constrained form:

$$\min_x f(x) \quad \text{subject to} \quad h(x) \leq t$$

Lagrange form:

$$\min_x f(x) + \lambda h(x)$$

Show that they are Equivalent?

Solution
First of all, they are not equivalent in general. A counterexample: If we let $f(x) = x, h(x) = -x$ and $t = 1$, then there exists no λ that can help the Lagrange form to yield the same solution.

Given a t, what should λ be such that the Lagrange form yields the same solution as the constrained form?

We have to impose a few extra conditions: (1) both $f(x)$ and $h(x)$ are at least twice differentiable; (2) the function $g(x) := -f'(x)/h'(x)$ is well-defined and strictly monotonic; (3) the function $j(x) = f''(x) + g(x)h''(x)$ is well-defined and strictly positive.
For any given t, let's denote the optimal solution to the constrained problem as $\tilde{x}(t)$. We claim that when

$$\lambda(t) = -f'[\tilde{x}(t)]/h'[\tilde{x}(t)]$$

the Lagrange form problem is equivalent to the constrained form problem. Taking the derivative of the Lagrangian $L(x) = f(x) - \frac{f'[\tilde{x}(t)]}{h'[\tilde{x}(t)]}h(x)$ and setting it to 0 gives,

$$f'(x) - \frac{f'[\tilde{x}(t)]}{h'[\tilde{x}(t)]}h'(x) = 0$$

Since the function $-f'(x)/h'(x)$ is strictly monotonic, the only possible solution is given by $x = \tilde{x}(t)$. Indeed, $x = \tilde{x}(t)$ is the optimal solution to the Lagrange problem because $L''(\tilde{x}(t)) > 0$ by assumption (3).

Given a $\lambda > 0$, what should t be such that the constrained form yields the same solution as the Lagrange form?

Let's denote $x^*(\lambda)$ as the optimal solution to the Lagrange problem and $\tilde{x}(t)$ as the optimal solution to the constrained problem. We claim that

$$t(\lambda) = h[x^*(\lambda)]$$

The optimality of $x^*(\lambda)$ for the Lagrange problem implies that

$$\forall x \in \{x : h(x) \leq h[x^*(\lambda)]\}, f[x^*(\lambda)] + \lambda h[x^*(\lambda)] \leq f(x) + \lambda h(x) \leq f(x) + \lambda h[x^*(\lambda)]$$

which implies that $f[x^*(\lambda)] \leq f(x)$ for all feasible x in the constrained problem. So $x^*(\lambda)$ is a solution to the constrained problem as well.

```
M = csvread('HW11.csv',2);
n = size(M,1);
p = size(M,2)-1;
X = M(1:n,1:p);
Y = M(1:n,p+1);

% Hard margin SVM
HardMarginSVM = fitcsvm(X,Y);%,'BoxConstraint',0.02);
sv = HardMarginSVM.SupportVectors;
figure;
gscatter(X(:,1),X(:,2),Y);
hold on;
plot(sv(:,1),sv(:,2),'ko','MarkerSize',10);
legend('-1','1','Support Vector');
m = - HardMarginSVM.Beta(1)/HardMarginSVM.Beta(2);
b = - HardMarginSVM.Bias/HardMarginSVM.Beta(2);
refline(m,b);
margin = 1/HardMarginSVM.Beta(2);
l1 = refline(m,b+margin);
l2 = refline(m,b-margin);
l1.LineStyle = '--';
l2.LineStyle = '--';
hold off;

% nsv = size(sv,1);
% for i = 1:nsv
%     sv(i,:)
%     HardMarginSVM.Beta(1)*sv(i,1)+HardMarginSVM.Beta(2)*sv(i,2)+HardMarginSVM.Bias
% end
% HardMarginSVM.Beta
% HardMarginSVM.Bias
```

```
L = load('MNIST_20x20.mat');
labels = L.labels;
imgs = L.imgs;
n = size(labels,1);
counter = zeros(4);
for i = 1:n
    if labels(i) == 1
        counter(1) = counter(1) + 1;
        IMG(counter(1),1) = i;
    elseif labels(i) == 2
        counter(2) = counter(2) + 1;
        IMG(counter(2),2) = i;
    elseif labels(i) == 3
        counter(3) = counter(3) + 1;
        IMG(counter(3),3) = i;
    elseif labels(i) == 4
        counter(4) = counter(4) + 1;
        IMG(counter(4),4) = i;
    end
end

% pair: 3 & 4
k = 3;
l = 4;
total = counter(k)+counter(l);
X = zeros(400,total);
Y = (total);
for i = 1:counter(k)
    X(1:400,i) = reshape(imgs(:,:,IMG(i,k)),[400,1]);
    Y(i) = 1;
end
for i = 1:counter(l)
    j = counter(k)+i;
    X(1:400,j) = reshape(imgs(:,:,IMG(i,l)),[400,1]);
    Y(j) = -1;
end
X = transpose(X);

PC = pca(X);
mean = ones(1,total)*X/total;
proj = zeros(total,2);
for i = 1:total
    proj(i,1) = (X(i,:)-mean)*PC(:,1);
    proj(i,2) = (X(i,:)-mean)*PC(:,2);
end

% % Hard SVM
% SVM_hard = fitcsvm(X,Y,'BoxConstraint',Inf);
% sv = SVM_hard.SupportVectors;
% figure;
% gscatter(proj(:,1),proj(:,2),Y);
% hold on;
% % plot(sv(:,1),sv(:,2),'ko','MarkerSize',10);
% % legend('-1','1','Support Vector');
% new_beta = SVM_hard.Beta'*PC(:,1:2);
% m = - new_beta(1)/new_beta(2);
% b = - SVM_hard.Bias/new_beta(2);
% refline(m,b);
```

```
% margin = 1/new_beta(2);
% l1 = refline(m,b+margin);
% l2 = refline(m,b-margin);
% l1.LineStyle = '--';
% l2.LineStyle = '--';
% hold off;
% count = 0;
% for i = 1:total
%     crit = X(i,:)*SVM_hard.Beta(:,1)+ SVM_hard.Bias;
%     if Y(i)*crit > 0
%         count = count + 1;
%     end
% end
% accuracy = count/total;

% % Soft SVM
% for i = 1:4
%     BC = 0.04+(i-1)*0.005;
%     SVM_soft = fitcsvm(X,Y,'BoxConstraint',BC);
%     CVSVM_soft = crossval(SVM_soft,'KFold',2);
%     Loss_soft = kfoldLoss(CVSVM_soft,'LossFun','classiferror') % average loss
% end
% Best BC = 0.04
% SVM_soft = fitcsvm(X,Y,'BoxConstraint',0.04);
% sv = SVM_soft.SupportVectors;
% figure;
% gscatter(proj(:,1),proj(:,2),Y);
% hold on;
% % plot(sv(:,1),sv(:,2),'ko','MarkerSize',10);
% % legend('-1','1','Support Vector');
% new_beta = SVM_soft.Beta'*PC(:,1:2);
% m = - new_beta(1)/new_beta(2);
% b = - SVM_soft.Bias/new_beta(2);
% refline(m,b);
% margin = 1/new_beta(2);
% l1 = refline(m,b+margin);
% l2 = refline(m,b-margin);
% l1.LineStyle = '--';
% l2.LineStyle = '--';
% hold off;
% count = 0;
% for i = 1:total
%     crit = X(i,:)*SVM_soft.Beta(:,1)+ SVM_soft.Bias;
%     if Y(i)*crit > 0
%         count = count + 1;
%     end
% end
% accuracy = count/total;

% SVM with Guassian kernel
SVM_Gaussian = fitcsvm(proj,Y,'BoxConstraint',↙
Inf,'KernelFunction','rbf','KernelScale','auto');
sv = SVM_Gaussian.SupportVectors;
figure;
gscatter(proj(:,1),proj(:,2),Y);
hold off;
% plot(sv(:,1),sv(:,2),'ko','MarkerSize',10);
% legend('-1','1','Support Vector');
```

```
% d = 2;
x1Grid = linspace(-6,7,27);
x2Grid = linspace(-6,7,27);
% [x1Grid,x2Grid] = meshgrid(min(proj(:,1)):d:max(proj(:,1)),min(proj(:,2)):d:max(proj(:,↙
2)));
% XGrid = [x1Grid(:),x2Grid(:)];
% [~,score] = predict(SVM_Gaussian,XGrid);
% % contour(x1Grid,x2Grid,reshape(score(:,2),size(x1Grid)),[0 0],'k');
% figure;
% gscatter(x1Grid,x2Grid,reshape(score(:,2),size(x1Grid)));
count = 0;
xx = zeros(size(x1Grid,2)*size(x2Grid,2),2);
p = zeros(size(x1Grid,2)*size(x2Grid,2),1);
for i = 1:size(x1Grid,2)
    for j = 1:size(x2Grid,2)
        count = count + 1;
        xx(count,:) = [x1Grid(i),x2Grid(j)];
        p(count,1) = predict(SVM_Gaussian,xx(count,:));
    end
end
figure;
gscatter(xx(:,1),xx(:,2),p(:,1));

crit = zeros(total);
count = 0;
for i = 1:total
    [~,crit(i)] = predict(SVM_Gaussian,proj(i,:));
    if Y(i)*crit(i) > 0
        count = count + 1;
    end
end
accuracy = count/total;
```