Università della Svizzera italiana
Faculty of Informatics

May 15, 2018

Stochastic Methods

Assignment 4

Sumeet Gyanchandani

1 Support Vector Machines

Listing 1. mysvm.m

```
clear all;
close all;

%load data
load("svm_data.mat");

xd = data(:,1);
yd = data(:,2);
zd = data(:,3);
class = data(:,4);

%1a 3D case
%visual data
pointsize = 30;
figure;
scatter3(xd,yd,zd,pointsize,class,'filled');
xlabel('X-axis');
ylabel('Y-axis');
zlabel('Z-axis');
hold on;

%Quadratic Programming
d = size(data,1);
n = size(data,2)-1; %not counting the class column

X = data(:,1:n);
Y = data(:,n+1);

H = eye(n+1);
H(n+1,n+1)=0; %To make sure we do not minimize constant b
f = zeros(n+1,1);

Z = [X ones(d,1)];
A = -diag(Y) * Z;
c = -1*ones(d,1);

[w, fval] = quadprog(H,f,A,c);
```

```
%Plotting Hyperplane
[X1, Y1] = meshgrid(-8:0.5:6,-10:10);
w1=w(1,1);
w2=w(2,1);
w3=w(3,1);
b=w(4,1);
C1 = zeros(21,29);
size(C1);
Z1=-(w1*X1+w2*Y1-b)/w3; %Optimal hyperplane
surf(X1,Y1,Z1,C1);
view(40,10);
hold off;

%1b 2D case
%visual data
figure;
scatter(xd, yd, pointsize, class, 'filled');
hold on;

%Quadratic Programming
d = size(data,1);
n = size(data,2)-2; %also not counting the 3rd dimension

X = data(:,1:n);
Y = data(:,n+2);

H = [eye(n), zeros(n,d+1); zeros(d+1,n), zeros(d+1,d+1)];
f = [zeros(n+1,1);.3*ones(d,1)];
leq = [-inf;-inf;-inf;zeros(d,1)];

c = -ones(d,1);
Z = [X -c -eye(d,d)];
A = -diag(Y) * Z;


options = optimoptions('quadprog','Algorithm','interior-point-convex','Display','off');
w = quadprog(H,f,A,c,[],[],leq,[],[],options);

%Plotting
X2 = [-8:0.5:6];
Y2 = (-X2*w(1) - w(3))/w(2);
xlim([-8 6])
ylim([-10 10])
plot(X2,Y2,'k-');
hold off;
```

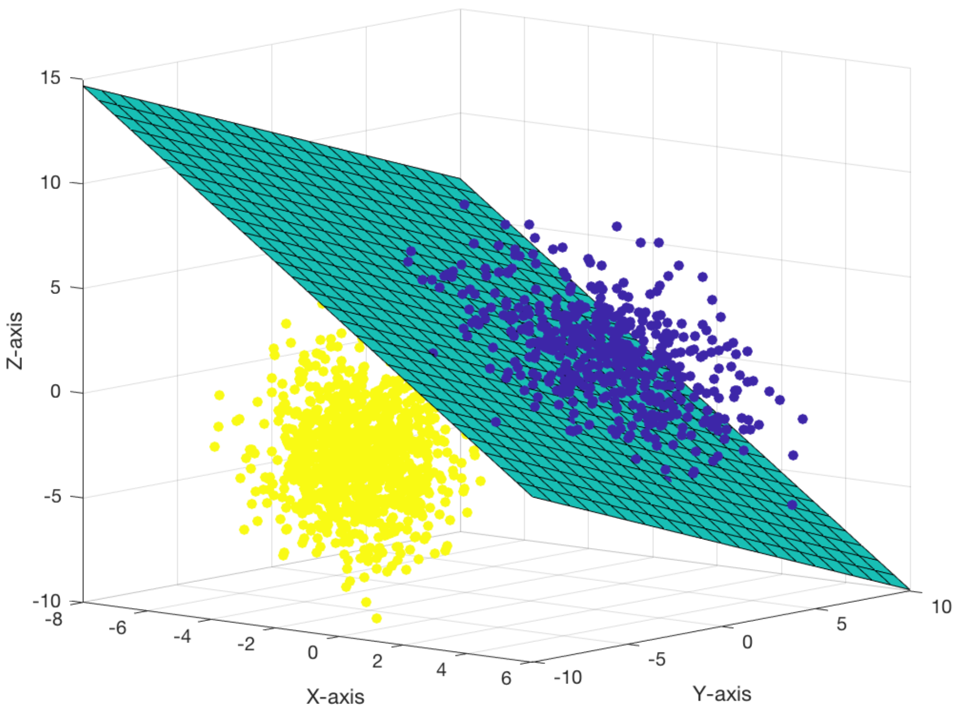

Figure 1. Hyperplane in 3D space

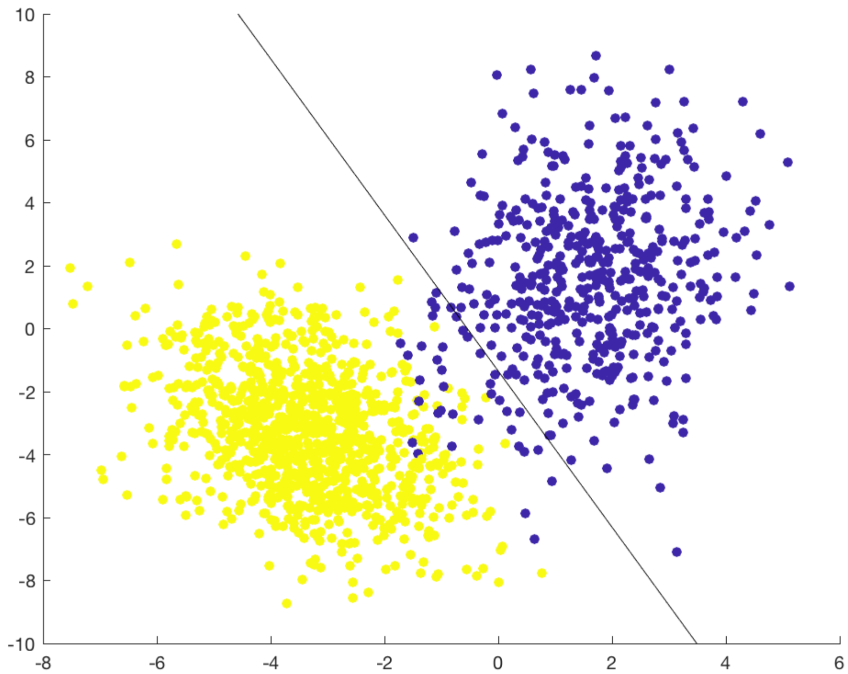

Figure 2. Hyperplane in 2D space

2 AIC for Linear Regression

The sequence of one dimensional response variables $Y_1, \ldots, Y_n$ and a sequence of covariates $X_1, \ldots, X_n$ with $X_i = [x_{i1}, \ldots, x_{ip}]$ for $i = 1, ..., n$. Linear Regression is used to describe the dependence between Y and X:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i, \text{ with } \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Y is Gaussian Distributed, hence

$$Y_i = \mathcal{N}(\mu_i, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{(x_i-\mu_i)^2}{2\sigma^2}}$$

therefore the likelihood function is

$$L(\mu_i, \sigma^2)) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu_i)^2}{2\sigma^2}}$$

and the log-likelihood function is

$$= \log((2\pi\sigma^2)^{-\frac{n}{2}}) + \log(e^{-\frac{\sum_{i=1}^{n}(x_i-\mu_i)^2}{2\sigma^2}}) \tag{1}$$

$$= \frac{n}{2}[-\log(\sigma^2) - \log(2\pi) + 1] \tag{2}$$

Now, we know that

$$\text{AIC} = -2LL + 2k$$

where, k is the number of parameters. We have $p + 1$ parameters for $\beta_0, \ldots, \beta_p$ and 1 parameter for the σ. Hence we have a total of $k = p + 2$ parameters. Substituting the value of k and equation 2 in the AIC formula:

$$\begin{aligned}\text{AIC}(p) &= -2\ (\frac{n}{2}[-\log(\sigma^2) - \log(2\pi) + 1]) + 2\ (p + 2)\\ &= n\log(\sigma^2) + n\ (1 + \log(2\pi)) + 2\ (p + 2)\end{aligned}$$