Test shows 282 errors (i.e. 10.84% misclassification rate). For sanity check, majority class are 0s (2788 samples). Misclassification rate for majority class is 0.

Code for quantization

|  |
| --- |
| function x\_quant=quant(x,med)  x=x-repmat(med,[size(x,1) 1]);  for i=1:size(x,1)  for j=1:size(x,2)  if x(i,j)<0 || x(i,j)==0  x\_quant(i,j)=1;  else  x\_quant(i,j)=2;  end  end  end  end |

Code for classifier

|  |
| --- |
| function y\_pred=fhat(x\_test)    global pihat ghat\_01 ghat\_02 ghat\_11 ghat\_12    g1=pihat;  for j=1:numel(x\_test)  if x\_test(j)==2  g1=g1\*ghat\_12(j);  else  g1=g1\*ghat\_11(j);  end  end    g0=1-pihat;    for j=1:numel(x\_test)  if x\_test(j)==2  g0=g0\*ghat\_02(j);  else  g0=g0\*ghat\_01(j);  end  end    if g1>g0 || g1==g0  y\_pred=1;  else  y\_pred=0;  end  end |

Code for determining majority class

|  |
| --- |
| close all; clear all; clc  z=dlmread('spambase.data',',');  rng(0);  rp=randperm(size(z,1));  z=z(rp,:);  x=z(:,1:end-1);  y=z(:,end);  med=median(x);  x=quant(x,med);  ind\_nonzero=find(y);  numel(ind\_nonzero)  ind\_zero=find(y-1);  numel(ind\_zero) |

Code for generating results

|  |
| --- |
| close all; clear all; clc  global pihat ghat\_01 ghat\_02 ghat\_11 ghat\_12  z=dlmread('spambase.data',',');  rng(0);  rp=randperm(size(z,1));  z=z(rp,:);  x=z(:,1:end-1);  y=z(:,end);  med=median(x);  x=quant(x,med);  x\_train=x(1:2000,:);  y\_train=y(1:2000);  x\_test=x(2001:end,:);  y\_test=y(2001:end);  n=size(y\_train,1);  n\_1=sum(y\_train);  pihat=n\_1/n;  for j=1:size(x\_train,2)  sum=0;  for i=1:size(x\_train,1)  if x\_train(i,j)==2 && y\_train(i)==1  sum=sum+1;  end  end  n\_12(j)=sum;  ghat\_12(j)=n\_12(j)/n\_1;  ghat\_11(j)=1-ghat\_12(j);  end  for j=1:size(x\_train,2)  sum=0;  for i=1:size(x\_train,1)  if x\_train(i,j)==2 && y\_train(i)==0  sum=sum+1;  end  end  n\_02(j)=sum;  ghat\_02(j)=n\_02(j)/(n-n\_1);  ghat\_01(j)=1-ghat\_02(j);  end  for i=1:size(x\_test,1)  y\_pred(i)=fhat(x\_test(i,:));  end  y\_pred=y\_pred';  error=abs(y\_pred-y\_test);  errorSum=0;  for i=1:numel(error)  errorSum=errorSum+error(i);  end  errorSum  misclas\_rate=errorSum/numel(y\_test) |

Code for sanity check

|  |
| --- |
| close all; clear all; clc  global pihat ghat\_01 ghat\_02 ghat\_11 ghat\_12  z=dlmread('spambase.data',',');  rng(0);  rp=randperm(size(z,1));  z=z(rp,:);  x=z(:,1:end-1);  y=z(:,end);  med=median(x);  x=quant(x,med);  ind\_zero=find(y-1);  x\_train=x(ind\_zero,:);  y\_train=y(ind\_zero)  x\_test=x\_train;  y\_test=y\_train;  n=size(y\_train,1);  n\_1=sum(y\_train);  pihat=n\_1/n;  for j=1:size(x\_train,2)  sum=0;  for i=1:size(x\_train,1)  if x\_train(i,j)==2 && y\_train(i)==1  sum=sum+1;  end  end  n\_12(j)=sum;  ghat\_12(j)=n\_12(j)/n\_1;  ghat\_11(j)=1-ghat\_12(j);  end  for j=1:size(x\_train,2)  sum=0;  for i=1:size(x\_train,1)  if x\_train(i,j)==2 && y\_train(i)==0  sum=sum+1;  end  end  n\_02(j)=sum;  ghat\_02(j)=n\_02(j)/(n-n\_1);  ghat\_01(j)=1-ghat\_02(j);  end  for i=1:size(x\_test,1)  y\_pred(i)=fhat(x\_test(i,:));  end  y\_pred=y\_pred';  error=abs(y\_pred-y\_test);  errorSum=0;  for i=1:numel(error)  errorSum=errorSum+error(i);  end  misclas\_rate=errorSum/numel(y\_test) |