ECE 271A- HW2

Reya Sadhu
A59026753

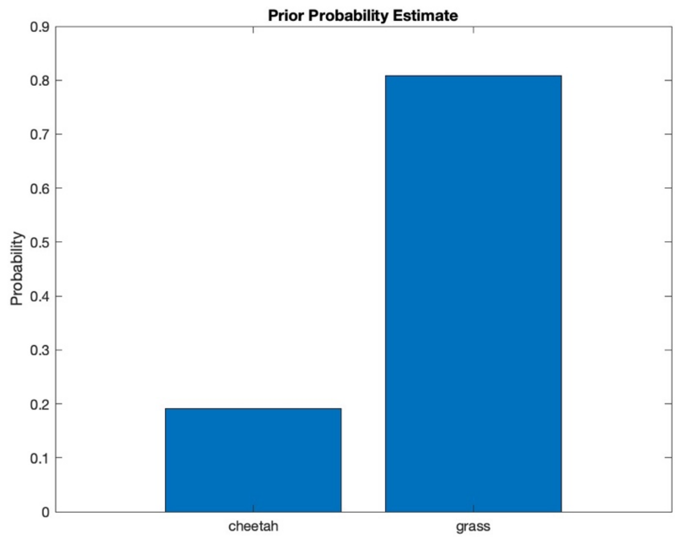
6.a) From problem 2, we find that the ML estimate of $P_X(k)$ is,

$$\pi_k^* = \frac{c_k}{n}$$

Where c_k is the no of observations in k-th class and n is total no of samples.
Using that, our prior probabilities are,

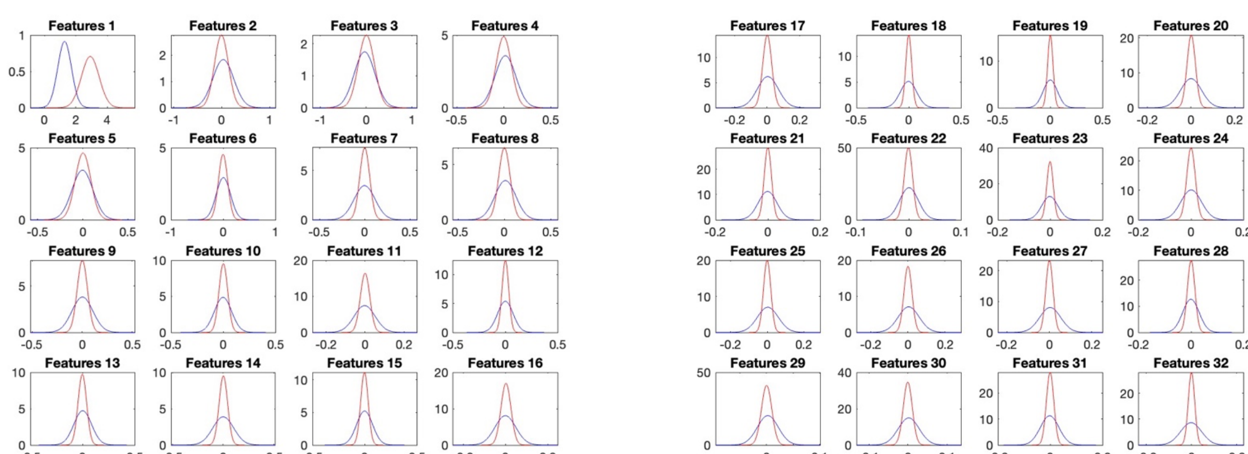
$$P_X(cheetah) = \frac{No\ of\ samples\ in\ cheetah\ class}{Total\ no\ of\ samples} = \frac{250}{250+1053} = .1919$$

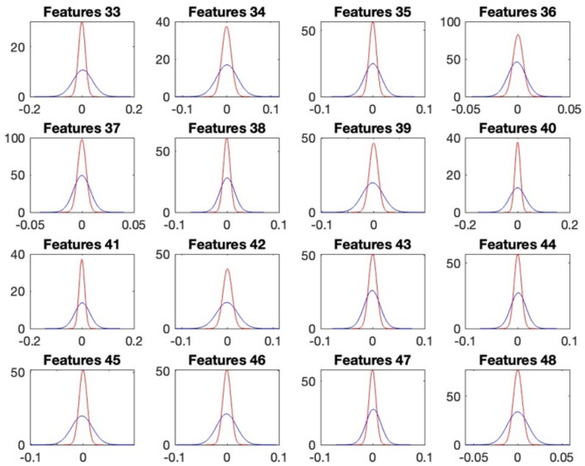
$$P_X(grass) = \frac{No\ of\ samples\ in\ grass\ class}{Total\ no\ of\ samples} = \frac{1053}{250+1053} = .8081$$

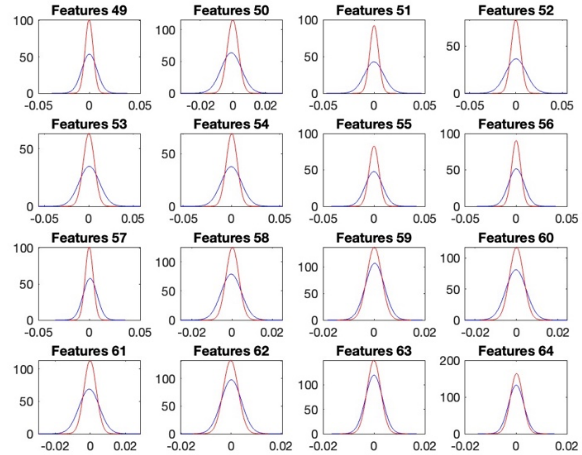


These are the same as the prior probabilities we computed in HW1. For this case, the intuitive estimate was same as our Maximum likelihood estimate.

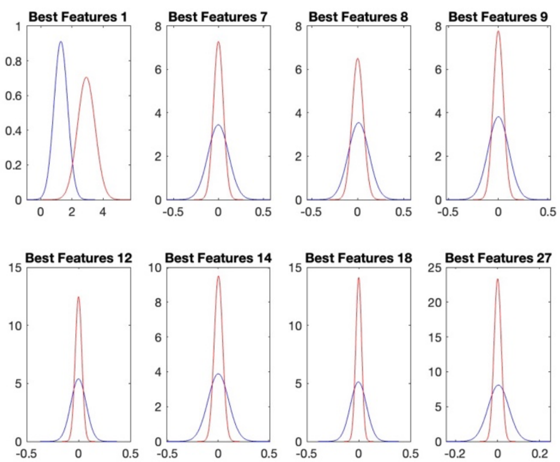
b) Assuming the 64 features in the DCT coefficients vector $X=\{X_1, X_2,\ldots, X_{64}\}$ are Gaussian, we can estimate their mean and variance from their sample mean and variance. Plotting the conditional probabilities $P_{X|Y}(X|cheetah)$ and $P_{X|Y}(X|grass)$ using these estimates, where the red line is for class cheetah and blue is for class grass.

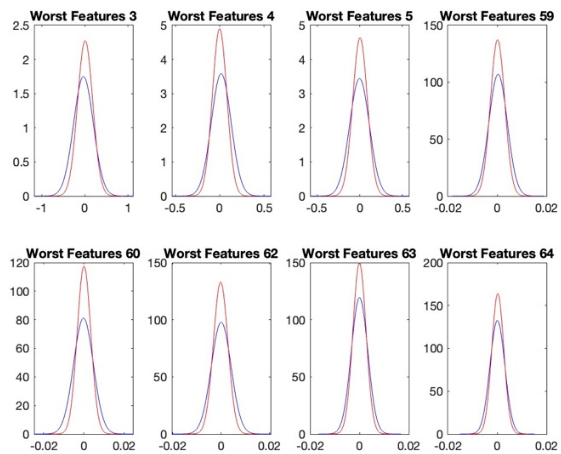






Now, the best features for the classification purpose will be where there is a considerable difference between $P_{X|Y}(X|cheetah)$ and $P_{X|Y}(X|grass)$ for all x's. Except feature 1, all the other features overlap each other with an almost similar mean. So, we choose the distributions based on the spread(variance). By a visual inspection, we choose the best 8 features to be [1,7,8,9,12,14,18,27] and the worst 8 features to be [3,4,5,59,60,62,63,64]. By keeping the plots side by side , we can clearly see the difference.





For the best features, the two conditional distributions are clearly separated from each other, while in the worst features they are overlapping each other. Which means worst features are nearly same for both classes, and thus not reliable for classification purpose.

c) The Bayesian Decision Rule says,

$$i^*(x) = \underset{i}{arg\,max}\; g_i(x) = \underset{i}{arg\,max}\; P_{i|X}(i|x) = \underset{i}{arg\,max}\; P_{X|i}(x|i)P(i)$$

Here, the X is a vector. So we can write ,

$$P_{X|i}(x|i) = \frac{1}{\sqrt{(2\pi)^d|\Sigma_i|}}\, exp\,\{-\frac{1}{2}(x-\mu_i)^T{\Sigma_i}^{-1}(x-\mu_i)\}$$

Where μ_i is the mean vector for the features in ith class and Σ_i is the covariance matrix of the features when in ith class.

As there are only two classes, the posterior probabilities can be written as,

$$P_{i|X}(i = cheetah|x) = \frac{P_{X|i}(x|i = cheetah) * P(cheetah)}{P_{X|i}(x|i = cheetah) * P(cheetah) + P_{X|i}(x|grass) * P(grass)}$$

So, the decision function can be written as,

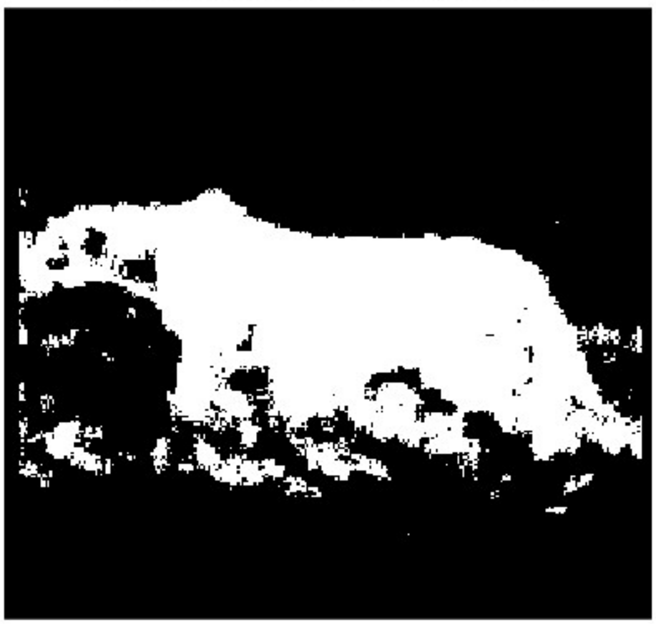
$$g_0(x) = \frac{1}{1 + expd_0(x - \mu_0) - d_1(x - \mu_1) + \alpha_0 - \alpha_1}$$

$$where, \qquad d_i(x, y) = (x - y)^T \Sigma_i^{-1} (x - y)$$

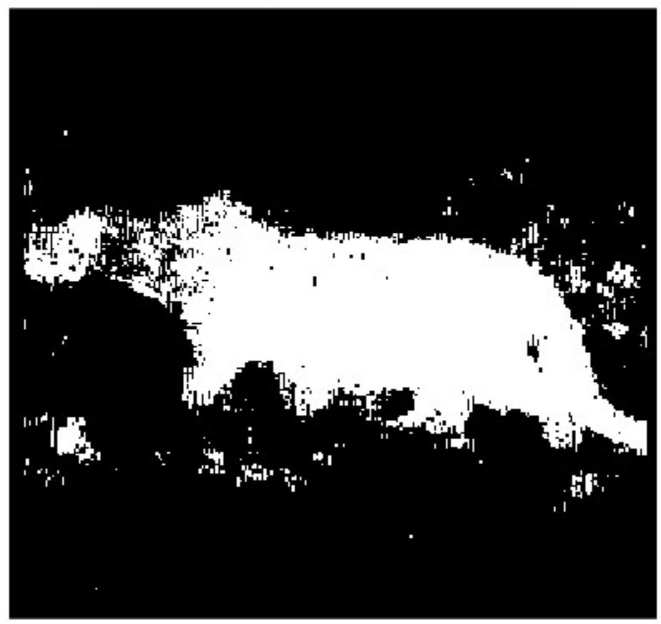
$$\alpha_i = log(2\pi)^d |\Sigma_i - 2logP(i)$$

And if $g_{cheetah}(x)>0.5$ then classify that pixel as cheetah.
Here, we are calculating the dct by taking 8*8 blocks around each pixel, so I assigned the predicted class to the center pixel of the block. Then we pad the mask symmetrically to match the dimension of input image and we get this result.

Mask with all 64 features



Mask with best 8 features



We calculate the probability or error by this formula,

$$P(error) = \int P_{i,x}(i \neq g^*(x), x)\ dx$$

This can be simplified into,

$$P(error) = \frac{No\ of\ mismatches}{Total\ number\ of\ pixels}$$

So, with 64 features, P(error) = 0.080523
With 8 best features, P(error) = 0.051736

We can see using the 8 best features gives us a better result than using all the 64 features. Its because all the 64 features include those features which have a similar density function and thus cannot distinguish the classes. So, they skew the classification decision and give more error. This is a case of high dimensionality, where we should only consider the useful dimensions rather than all to get a better accuracy.

Code:

```
load('/Users/reyasadhu/Downloads/trainingSamplesDCT_8_new.mat');

numCheetahSamples = size(TrainsampleDCT_FG, 1);
numGrassSamples = size(TrainsampleDCT_BG, 1);


totalSamples = numCheetahSamples + numGrassSamples;


PY_cheetah = numCheetahSamples / totalSamples;
PY_grass = numGrassSamples / totalSamples;

disp(['PY(cheetah) = ' num2str(PY_cheetah)]);
disp(['PY(grass) = ' num2str(PY_grass)]);

prior=[PY_cheetah,PY_grass];
figure;
bar(["cheetah";"grass"],prior);
title('Prior Probability Estimate');
ylabel('Probability');
ax = gca;
exportgraphics(ax,"/Users/reyasadhu/Desktop/Masters/ECE 271A Statistical
Learning/HW2/priors.jpg")

%As the features follow a Gaussian distribution, the sample mean and sample
%variance are the best ML estimators
mean_cheetah=mean(TrainsampleDCT_FG,1);
mean_grass=mean(TrainsampleDCT_BG,1);

std_cheetah=std(TrainsampleDCT_FG,0,1);
std_grass=std(TrainsampleDCT_BG,0,1);

for i=1:64
    % Taking range of x between 5 times the std deviation from the mean
    x_cheetah(i,:)=(mean_cheetah(i)-
5*std_cheetah(i)):std_cheetah(i)/100:(mean_cheetah(i)+5*std_cheetah(i));
    Px_cheetah(i,:)= normpdf(x_cheetah(i,:),mean_cheetah(i),std_cheetah(i));

    x_grass(i,:)=mean_grass(i)-
5*std_grass(i):std_grass(i)/100:mean_grass(i)+5*std_grass(i);
    Px_grass(i,:) = normpdf(x_grass(i,:),mean_grass(i),std_grass(i));
end


% Plotting P(xi|cheetah) and P(xi|grass) desnsities of for i=[1,64]
for k=0:3
    fig=figure;
    for i = 1:16
        subplot(4,4,i);
        plot(x_cheetah(i+16*k, :),Px_cheetah(i+16*k, :),'-b',x_grass(i+16*k,
:),Px_grass(i+16*k, :),'-r');
        title(['Features ',num2str(i+16*k)]);
    end
    print(fig,'-djpeg',sprintf("/Users/reyasadhu/Desktop/Masters/ECE 271A Statistical
Learning/HW2/All_features %d.jpg",k+1));
end

best=[1,7,8,9,12,14,18,27];
worst=[3,4,5,59,60,62,63,64];
```

```
fig=figure;
for i=1:8
    subplot(2,4,i);
    ix=best(i);
    plot(x_cheetah(ix, :),Px_cheetah(ix, :),'-b',x_grass(ix, :),Px_grass(ix, :),'-r');
    title(['Best Features ',num2str(ix)]);
end
print(fig,'-djpeg',"/Users/reyasadhu/Desktop/Masters/ECE 271A Statistical 
Learning/HW2/Best_features.jpg");

fig=figure;
for i=1:8
    subplot(2,4,i);
    ix=worst(i);
    plot(x_cheetah(ix, :),Px_cheetah(ix, :),'-b',x_grass(ix, :),Px_grass(ix, :),'-r');
    title(['Worst Features ',num2str(ix)]);
end
print(fig,'-djpeg',"/Users/reyasadhu/Desktop/Masters/ECE 271A Statistical 
Learning/HW2/Worst_features.jpg");

% calculating covariance matrix and alpha values for the training samples
cov_cheetah_64=cov(TrainsampleDCT_FG);
cov_grass_64=cov(TrainsampleDCT_BG);
alpha_cheetah_64=log(((2*pi)^64)*det(cov_cheetah_64))-2*log(PY_cheetah);
alpha_grass_64=log(((2*pi)^64)*det(cov_grass_64))-2*log(PY_grass);


dct_fg_8=TrainsampleDCT_FG(:,best);
dct_bg_8=TrainsampleDCT_BG(:,best);

mean_cheetah_best8=mean_cheetah(best);
mean_grass_best8=mean_grass(best);

cov_cheetah_best8=cov(dct_fg_8);
cov_grass_best8=cov(dct_bg_8);
alpha_cheetah_best8=log(((2*pi)^8)*det(cov_cheetah_best8))-2*log(PY_cheetah);
alpha_grass_best8=log(((2*pi)^8)*det(cov_grass_best8))-2*log(PY_grass);


I = imread('/Users/reyasadhu/Downloads/homework1/cheetah.bmp');
I=im2double(I);


A = zeros(size(I, 1) - 7,size(I, 2) - 7);
A_best8 = zeros(size(I, 1) - 7,size(I, 2) - 7);

for i = 1:size(I, 1) - 7
    for j = 1:size(I, 2) - 7
        block = I(i:i+7, j:j+7);
        block_dct = dct2(block);
        dct_flat=zigzaged(block_dct);
        dct_flat_best8=dct_flat(best);
        g_cheetah=1/(1+exp(dxy(dct_flat,mean_cheetah,cov_cheetah_64)-
dxy(dct_flat,mean_grass,cov_grass_64)+alpha_cheetah_64-alpha_grass_64));
        if g_cheetah>0.5
            A(i+3,j+3)=1
        end

g_cheetah_best8=1/(1+exp(dxy(dct_flat_best8,mean_cheetah_best8,cov_cheetah_best8)-
```

```
dxy(dct_flat_best8,mean_grass_best8,cov_grass_best8)+alpha_cheetah_best8-
alpha_grass_best8));
        if g_cheetah_best8>0.5
            A_best8(i+3,j+3)=1
        end
    end
end
% Padding the image with zeros
A_resized=zeros(255,270);
A_best8_resized=zeros(255,270);
for i=4:251
    for j=4:266
        A_resized(i,j)=A(i-3,j-3);
        A_best8_resized(i,j)=A_best8(i-3,j-3);
    end
end

figure;
subplot(1,2,1);
imshow(A_resized);
title("Mask with all 64 features");
ax = gca;
exportgraphics(ax,"/Users/reyasadhu/Desktop/Masters/ECE 271A Statistical
Learning/HW2/mask_64.jpg");
subplot(1,2,2);
imshow(A_best8_resized);
title("Mask with best 8 features");
ax = gca;
exportgraphics(ax,"/Users/reyasadhu/Desktop/Masters/ECE 271A Statistical
Learning/HW2/mask_best8.jpg");


im_test = imread('/Users/reyasadhu/Downloads/homework1/cheetah_mask.bmp');
im_test=im2double(im_test);
err=abs(im_test-A_resized);
prob_err=sum(err,"all")/(255*270);
disp(['Probability error with 64 features:' num2str(prob_err)]);
err2=abs(im_test-A_best8_resized);
prob_err2=sum(err2,"all")/(255*270);
disp(['Probability error with 8 best features:' num2str(prob_err2)]);

imwrite(A_resized,"/Users/reyasadhu/Desktop/Masters/ECE 271A Statistical
Learning/HW2/mask_64.bmp");
imwrite(A_best8_resized,"/Users/reyasadhu/Desktop/Masters/ECE 271A Statistical
Learning/HW2/mask_best8.bmp");

function output= zigzaged(input)
    zigzag=importdata('/Users/reyasadhu/Downloads/homework1/Zig-Zag Pattern.txt') ;
    zigzag=zigzag+1;
    output=zeros(1,64);
    for i=1:8
        for j=1:8
            output(zigzag(i,j))=input(i,j);
        end
    end
end

function output=dxy(x,y,cov)
    output=(x-y)*inv(cov)*transpose(x-y);
end
```