Table of Contents


```
clc; clear;
```

Part 1

In this part, we were first asked to fix the gaussfit function. We are then asked to load the employees.mat file and use gaussfit to better understand the salaries of different departments.

1.1

First we show that gaussfit works. I assume gaussfit returns [mu, sigma^2] where mu = mean, sigma^2 = variance. There was conflicting information in MATLAB vs PDF assignment as to which one to return, and returning the variance maintains higher numerical precision. We see that the output matches what we expect. We expected to find that mu = [1 2 3 4 5] and sigma = 3. We get pretty close :)

```
mu_s = [1 2 3 4 5 6 7];
sigma_s = 3 * eye(size(mu_s, 2));
n = 1000000;
X = mvnrnd(mu_s,sigma_s,n)';  % n random samples from multivariate
 gaussian
[mu, sigma] = gaussfit(X)
```

```
mu =

    0.9978
    2.0013
    2.9993
    3.9941
    5.0022
    6.0000
    6.9988


sigma =
```

```
    3.0005
```

1.2, 1.3

Next we load in the employee data

```
load("employees.mat");

max_dept = max(dept(:));  % get the maximum department number for
 iterating
mu = zeros(max_dept,1); % initialize matrices
sigma = zeros(max_dept,1);
for i = 1:max_dept
    [mu(i),sigma(i)] = gaussfit( sal(dept == i) );
end
```

1.4

Now we find the departments that have the highest and lowest mean salary

```
[~, min_mu] = min(mu);
[~, max_mu] = max(mu);
deptnames = fieldnames(depts);  % store strings of names of the
 departments
deptnumbers = struct2cell(depts);
deptnumbers = [deptnumbers{:}];
disp("Dept with lowest mean = " + min_mu + " (" +
 deptnames(deptnumbers==min_mu) + "), mu = $" + round(mu(min_mu), 2));
disp("Dept with highest mean = " + max_mu + " (" +
 deptnames(deptnumbers==max_mu) + "), mu = $" + round(mu(max_mu), 2));
fprintf(1, '\n'); % empty line
```

```
Dept with lowest mean = 8 (FAMILY & SUPPORT), mu = $38635.3
Dept with highest mean = 29 (DoIT), mu = $94083.53
```

1.5

Now we find the departments that have the highest and lowest variance. This number is very big, but that is because its the variance, the standard deviation is much more reasonable ~42K. Also the min variance of 0 was a little bit suspicious, but in fact there is only one employee in that department, so that makes sense.

```
[~, min_s] = min(sigma);
[~, max_s] = max(sigma);
disp("Dept with lowest variance = " + min_s + " (" +
 deptnames(deptnumbers==min_s) + "), sigma^2 = $" + sigma(min_s));
disp("Dept with highest variance = " + max_s + " (" +
 deptnames(deptnumbers==max_s) + "), sigma^2 = $" + sigma(max_s));
```

```
Dept with lowest variance = 35 (LICENSE APPL COMM), sigma^2 = $0
```

```
Dept with highest variance = 14 (MAYOR'S OFFICE), sigma^2 =
 $1841837960.8433
```

Part 2

In this part we are supposed to complete the kernel density estimator using a Gaussian kernel. This function is in kde.m. We are going to use this density estimation with the crimes dataset in crimes.mat to discover areas with higher crime rates.

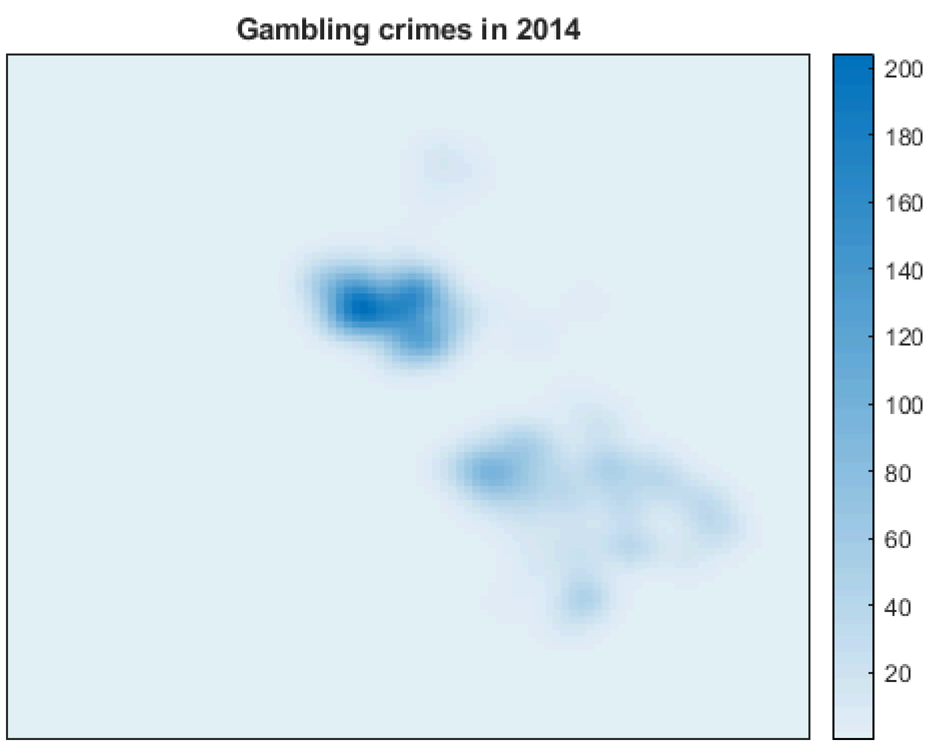
2.2, 2.3

Now we load in the crime data

```
load("crimes.mat")

% Now we look at gambling crimes in 2014
lats = lat(type == 15 & year == 2014);
lons = lon(type == 15 & year == 2014);
m = kdemap(lats, lons, 0.01, 100);
heat = heatmap(m, 'GridVisible', 'off', 'Title', "Gambling crimes in
 2014");

% plot formatting stuff
heat.XDisplayLabels = nan(size(heat.XDisplayData));
heat.YDisplayLabels = nan(size(heat.YDisplayData));
```

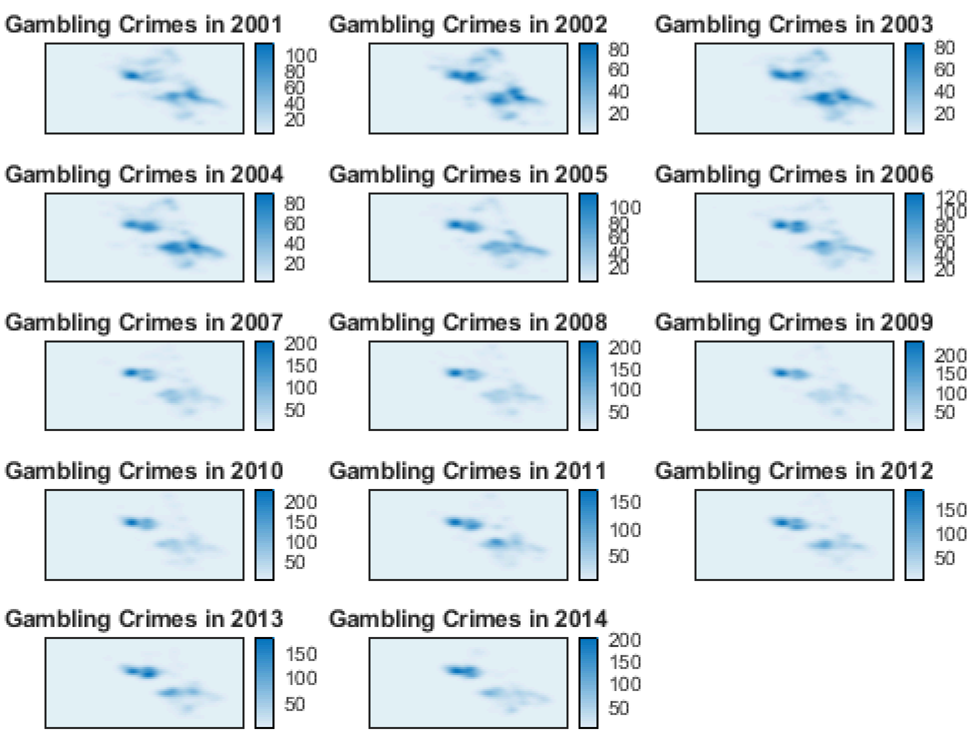


2.4

Now we show gambling crimes over a bunch of years! We show the gambling crimes from 2001 - 2014

```
figure()
tiledlayout(5,3)
for i = 2001:2014
    lats = lat(type == 15 & year == i);
    lons = lon(type == 15 & year == i);
    m = kdemap(lats, lons, 0.01, 100);
    nexttile
    heat = heatmap(m, 'GridVisible', 'off', 'Title', strcat("Gambling Crimes in ", num2str(i)));

    % plot formatting stuff
%     heat.GridVisible = 'off';
    heat.XDisplayLabels = nan(size(heat.XDisplayData));
    heat.YDisplayLabels = nan(size(heat.YDisplayData));
%     h.Title = strcat('Gambling Crimes in ', num2str(i));
end
```



2.5

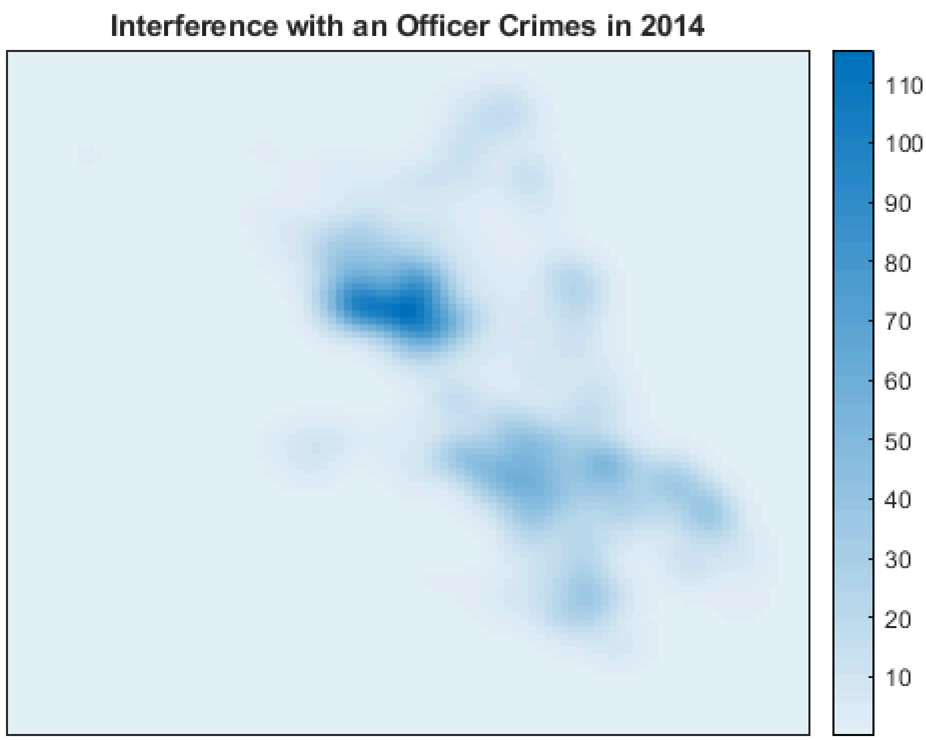
Based on these results, we can see that from 2001 to 2004 there were more widespread gambling crimes. By that I mean that throughout a larger area there was a higher chance of gambling crimes. There was also a much higher proportion of gambling crimes committed in the southern part of the region (as noted by the extra dark area in the heatmap) compared to recent years. We can also note that on the scale on the side of the heatmaps, gambling crimes seemed to be at a minimum from 2002 to 2004.

2.6

Next we show the crime for interference with an officer (crime 1) in 2014

```
lats = lat(type == 1 & year == 2014);
lons = lon(type == 1 & year == 2014);
m = kdemap(lats, lons, 0.01, 100);
figure()
heat = heatmap(m, 'GridVisible', 'off', 'Title', "Interference with an
 Officer Crimes in 2014");

% plot formatting stuff
heat.XDisplayLabels = nan(size(heat.XDisplayData));
heat.YDisplayLabels = nan(size(heat.YDisplayData));
```



Published with MATLAB® R2021a