Module 1

mean value

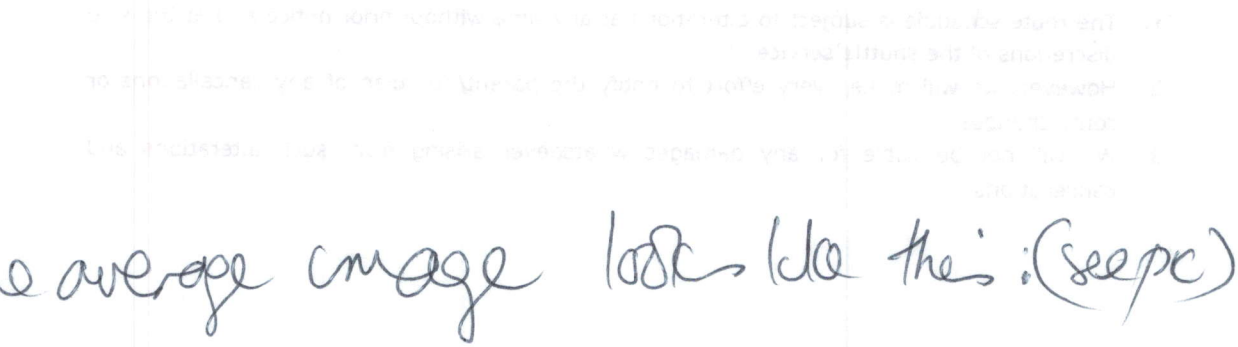
Data often compactly described by some of its statistical properties, such as mean and variance.

In this session, we will explain how to compute (various) means of data set

The mean of a data set, describe the average data point

The mean does not have to be a typical data point and it does not need to be part of the data set itself.

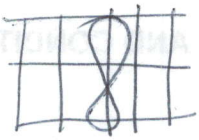
Eg if you look at a set of images of digits: (See pic below):



The average image looks like this: (see pic)

the properties of all images in dataset, but it is not part of the dataset itself. we obtain the average 8 as follows:

(see pic)



Remember, that an image can be represented easily as a long vector in a high dimensional vector space by stacking all pixels together (see pic):

after transforming all images into these vectors we take all image vectors in a dataset (see pic).

~~add~~

add them together, and divide by number of images in dataset.

This gives us the average image vector.
If we then reshape that vector into an image again
we get the average digit in dataset.

Here is an example with 4 8's. (see pic)

the mean of first one, is just the image itself., But
when we add the second image, we see that
the average image now contains properties of both
images

When we add the third image, the mean image,
is all three images on top of each other, divided by
three.
After the 4th image, we can still see characteristics
of all 4 images, in the average image (see pic)

when we add more images to our dataset, the average digit becomes more blurry (see pic)

and if we take all "8" images in our dataset, we get this 8 as the average image (see pic).

generally, if we have a dataset $x_1, \ldots x_n$:

$$\therefore \Delta = \{x_1, \ldots, x_n\}$$

we get the mean value, or expected value of the dataset as follows:

"Expected value of Δ, is one over number datapoints times sum, $n=1$ to N of x_n

$$E[\Delta] = \frac{1}{N} \sum_{n=1}^{N} x_n$$

∴ we sum up all datapoints in our dataset, and divide by number of datapoints (N) we have

Lets look at an example:

Create a dataset consisting of 5 numbers & get when & roll 5 dice: (see pic)

$D' = \{1, 2, 4, 6, 6\}$

lets call it D prime. The expected value of the average of this dataset, sum of all of elements in dataset, divided by number of elements in dataset

$$\therefore E[D'] = \frac{1+2+4+6+6}{6} = \frac{19}{5} = 3.8$$

we can clearly see that 3.8 is not part of dataset, and cannot ever be achieved by rolling a dice.

$\therefore$ its not a typical instance.

like we calculated the mean value of datasets, is the average datapoint.