

Module 1

mean values

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

In this session, we will explain how to compute (mean) means of data sets

The mean of a data set, describes 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.

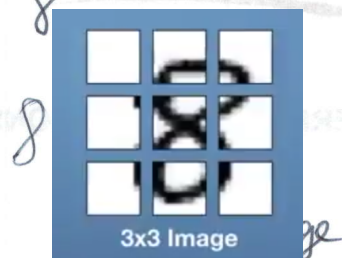
Eg if you look at a set of images of digit 8:
(See pic below):

8	8	8	8	8
8	8	8	8	8
8	8	8	8	8
8	8	8	8	8
8	8	8	8	8

The average image looks like this: (see pic)

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

(see pc)

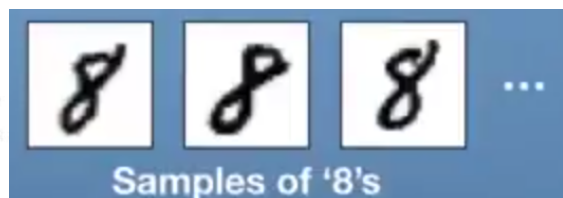


Remember, that an image can be represented as
a long vector in a high dimensional vector
space by stacking all pixels together (see pc):

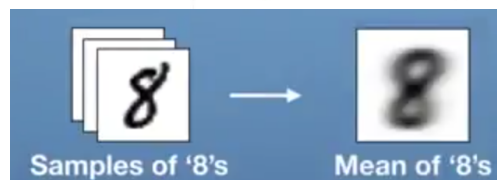


after transforming all images into the vectors
we take all image vectors in a dataset (see pc).

add



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



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 just 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 ~~and~~ 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, \dots, x_n :

$$\therefore D = \{x_1, \dots, x_n\}$$

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

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

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

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

Let's look at an example:

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

$$N = \{1, 2, 4, 6, 6\}$$



Let's call it Δ prime. The expected value of the average of the dataset, sum of all elements in dataset, divided by number of elements in dataset

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

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

\therefore it's not a typical instance.

Like we calculated the mean value of datasets, \hookrightarrow no average datapoint.