Dimensionality reduction techniques

Sanzhar Askaruly (San)

Ulsan National Institute of Science and Technology
Ph.D. Candidate in Biomedical Engineering

CodeSeoul MLA
December 10, 2022

Overview

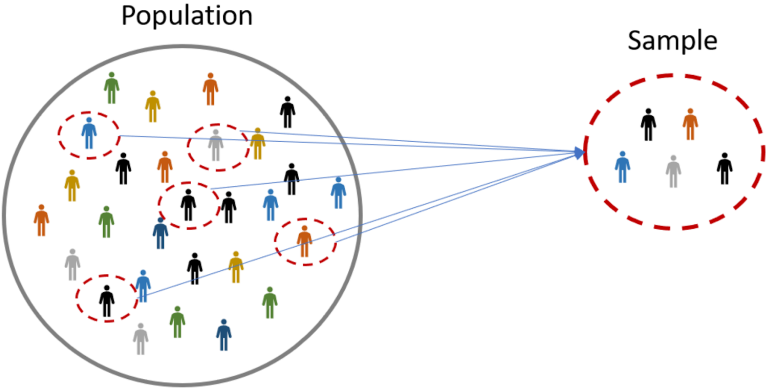
What we'll cover today:

1. Motivation
2. Background mathematics
   - Sample mean
   - Standard deviation
   - Variance
3. Examples
   - Iris dataset
4. Experiments
5. Summary
   - Conclusion
   - Practicum

Motivation

- Relationships between data

Population vs sample



Population vs sample

Population mean

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \tag{1}$$

N is number of items in the population

Sample mean

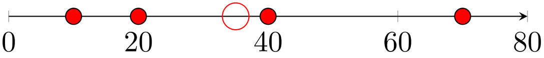
$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2}$$

n is number of items in the sample

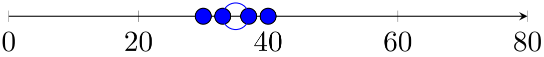
Standard deviation

Let's take a look on two samples:

$$A = [10\ 20\ 40\ 70]$$



$$B = [30\ 33\ 37\ 40]$$



Here, $\bar{A} = \bar{B} = 35$. Unfortunately, mean doesn't tell us a lot except for a middle point.

Standard deviation

For our two sets, $A = [10\ 20\ 40\ 70]$ and $B = [30\ 33\ 37\ 40]$, we would be more interested in the *spread* of the data. So, how do we calculate it?

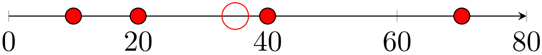
Standard deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)}} \tag{3}$$

In plain English, it is the "average distance from the mean of the data set to a point."

Standard deviation

Set 1: $A = [10\ 20\ 40\ 70]$, and $\bar{A} = 35$

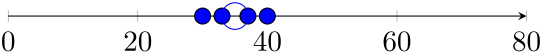


Let's calculate standard deviation:

A	$(A - \bar{A})$	$(A - \bar{A})^2$
10	-25	625
20	-15	225
40	5	25
70	35	1,225
Total		2,100
Divided by (n-1)		700
Square root		**26.4575**

Standard deviation

Set 2: $B = [30\ 33\ 37\ 40]$, and $\bar{B} = 35$



Let's calculate standard deviation:

B	$(B - \bar{B})$	$(B - \bar{B})^2$
30	-5	25
33	-2	4
37	2	4
40	5	25
Total		58
Divided by (n-1)		19.333
Square root		**4.397**

Variance

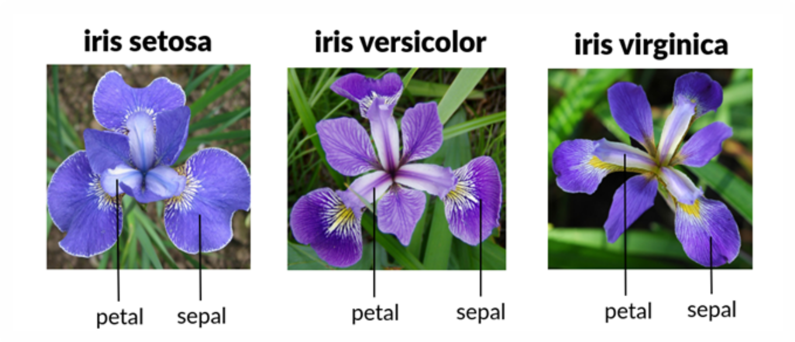
Similar to standard deviation So, how do we calculate it?

Standard deviation

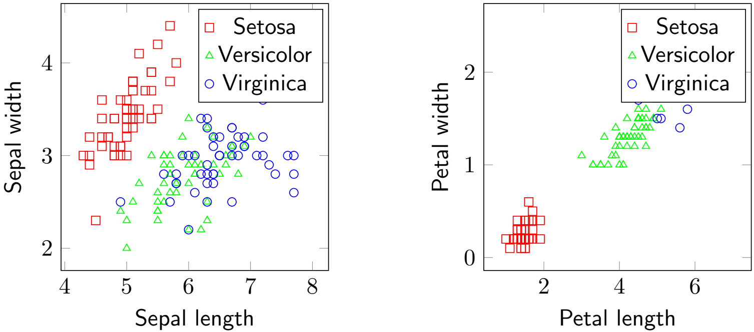
$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)}} \tag{4}$$

In plain English, it is the "average distance from the mean of the data set to a point."

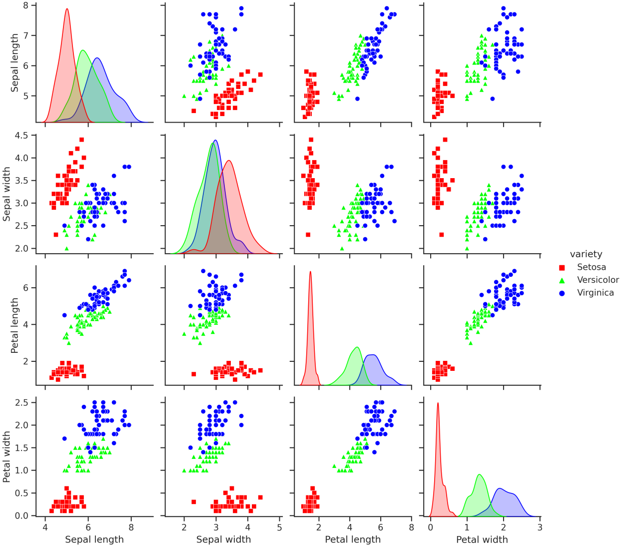
Iris flower dataset



Iris flower dataset



Iris flower dataset



Adagrad

Experiment

Experiment

Experiment

Experiment

Conclusion

Practicum

Thank you for your attention!

- Workshop contents:
  `https://github.com/CodeSeoul/machine-learning/tree/master/221210-pca`
- Follow-up QA?
  `http://discord.com/users/tuttelikz`

References

Ning Qian. "On the momentum term in gradient descent learning algorithms". In: *Neural networks* 12.1 (1999), pp. 145–151.

John Duchi, Elad Hazan, and Yoram Singer. "Adaptive subgradient methods for online learning and stochastic optimization.". In: *Journal of machine learning research* 12.7 (2011).

Sebastian Ruder. "An overview of gradient descent optimization algorithms". In: *arXiv preprint arXiv:1609.04747* (2016).

Mykola Novik. *torch-optimizer – collection of optimization algorithms for PyTorch.* Version 1.0.1. Jan. 2020.