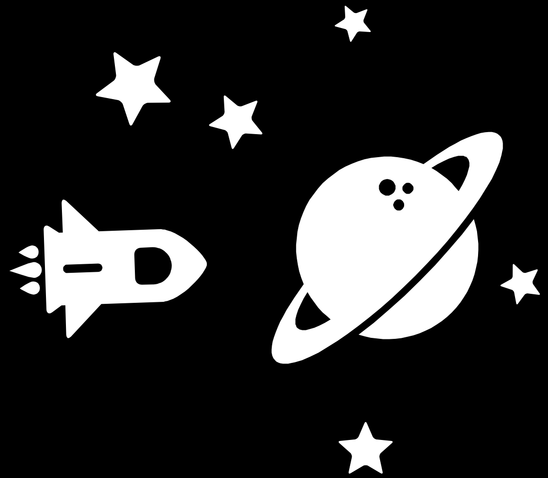
Feature Scaling:

Standardization & Normalization

Presentation by Berk Sudan

Benefits



Tunes Features (e.g, in distance-based algorithms)

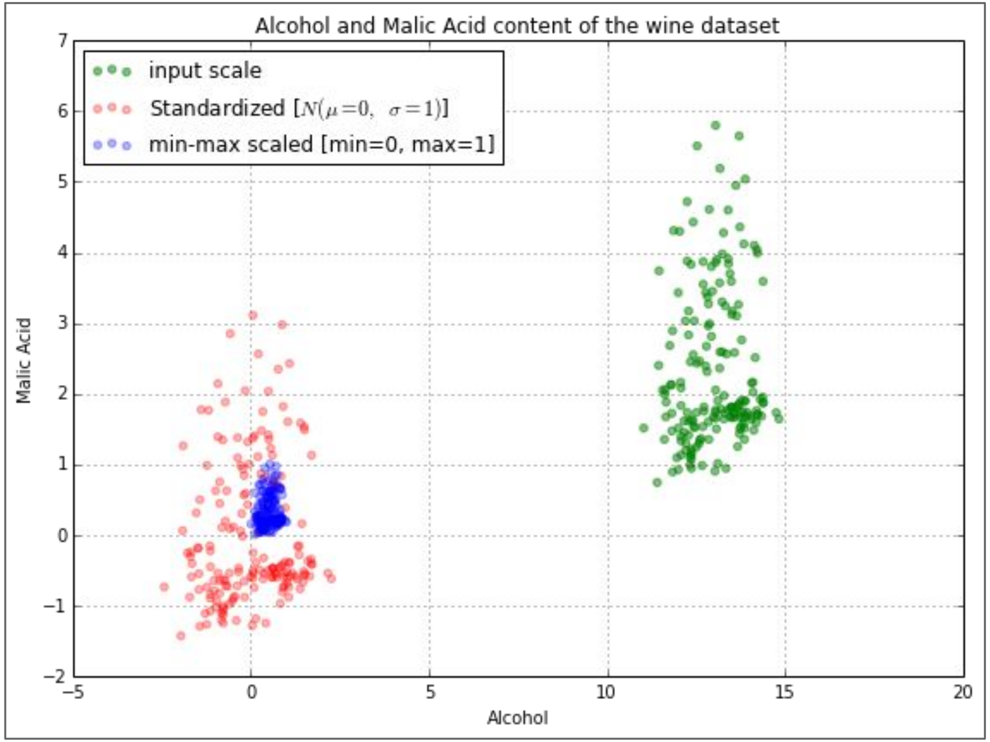
Age	Smoking Level	is Lung Cancer?
20	0	0
30	3	1
50	4	1

!!!!!

Avoiding Different Scales

	Class label	Alcohol	Malic acid
0	1	14.23	1.71
1	1	13.20	1.78
2	1	13.16	2.36
3	1	14.37	1.95
4	1	13.24	2.59

the features **Alcohol** (percent/volume) and **Malic acid** (g/L) are measured on different scales, so that ***Feature Scaling*** is necessary important prior to any comparison or combination of these data.



Ref: *https://sebastianraschka.com/Articles/2014_about_feature_scaling.html*

Good metrics (e.g, Silhouette for Clustering)

$$\boxed{a(i) = \frac{1}{|C_i| - 1} \sum_{j \in C_i, i \neq j} d(i, j)}$$

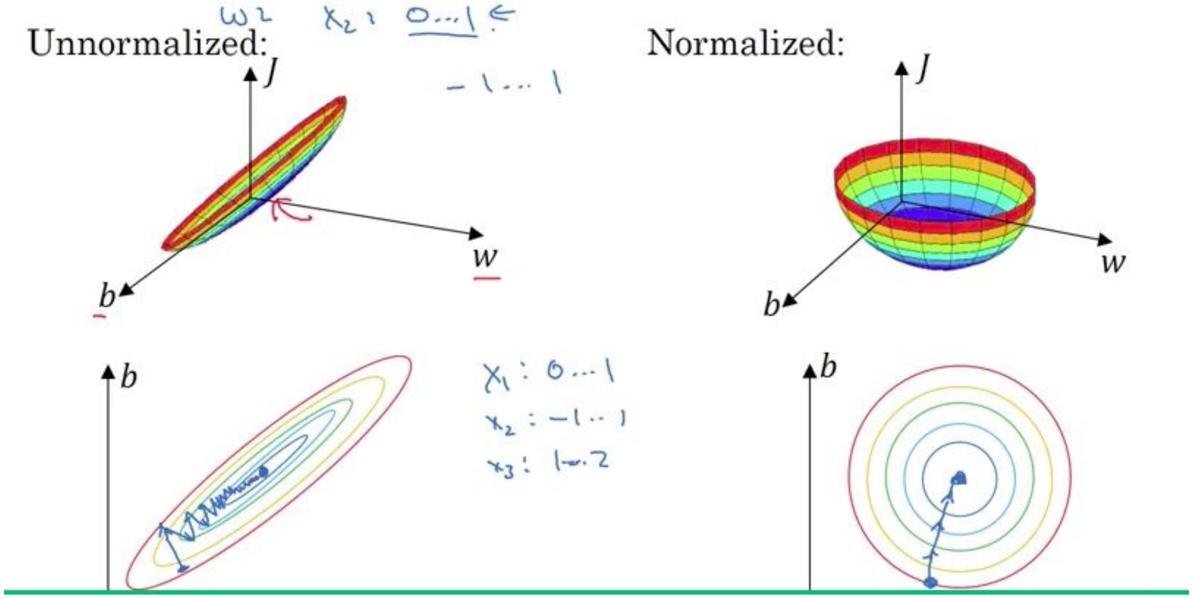
$$\boxed{b(i) = \min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} d(i, j)}$$

$$\boxed{s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}} \quad \sim (-1,1)$$

Ref: *https://en.wikipedia.org/wiki/Silhouette_(clustering)*

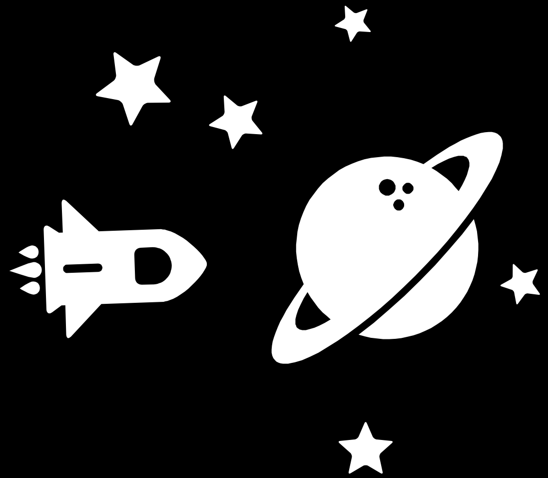
Convergence of Gradient Descent and Related Methods

logistic regression, **SVMs**, **perceptrons**, **neural networks** etc. if you are using g**radient descent/ascent-based optimization**, otherwise some weights will update **much faster than others.**

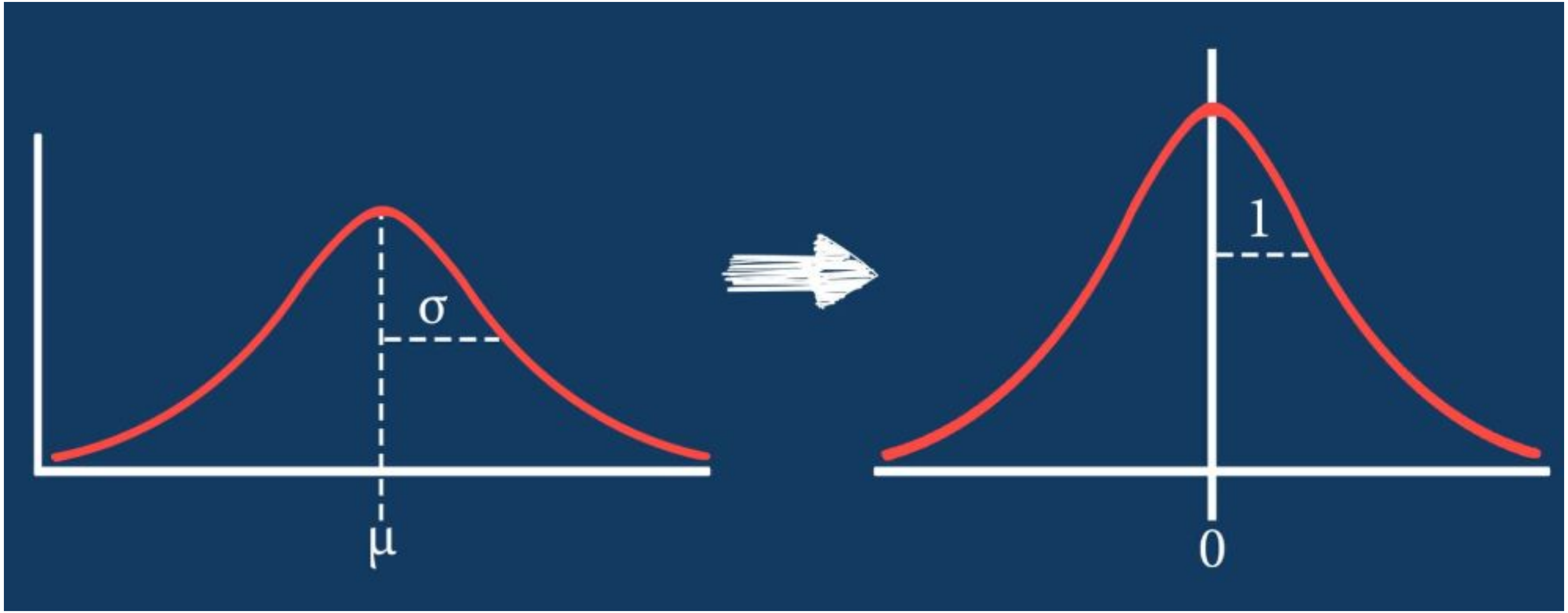


Ref: *Coursera Deep Learning Course 2 Week 1 notes*

Standardization



Gauss.. again!



Ref: *https://365datascience.com/standardization/*

Normal Distribution

$$N(\mu,\sigma^2)$$

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(\mu-x)^2}{2\sigma^2}}$$

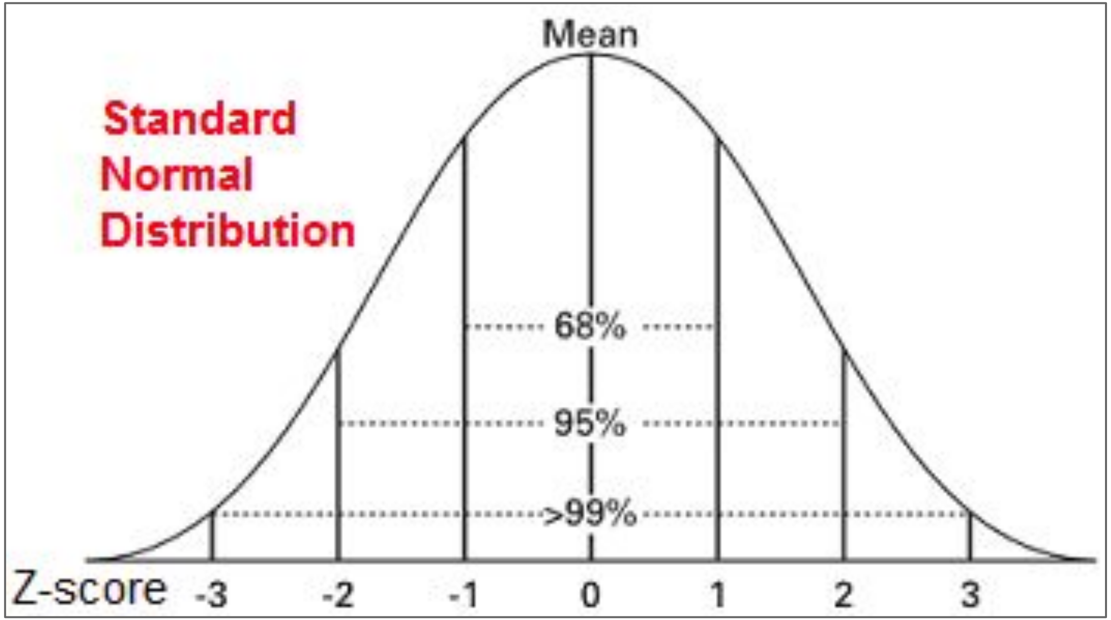
Standard Normal Distribution

$$N(0,1^2)$$

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$

Getting Z-Score

$$z = \frac{x - \mu}{\sigma}$$



Ref: *https://ecstep.com/wp-content/uploads/2016/10/standard-normal-distribution-6.png*

A Standardization Example

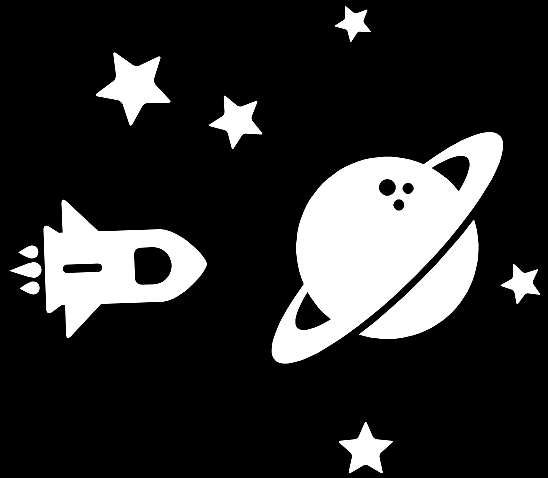
- Height of a pupil = 1.4m
- Mean for pupils = 1.2m
- STD for pupils = 0.4

-> z-score = (1.4-1.2) / 0.4 = **0.5**,

i.e. the pupil is **half** a **standard deviation** from the **mean**.

Ref: *https://365datascience.com/standardization/*

Normalization



Normalization Formula

$$v' = \frac{v - min_F}{max_F - min_F}$$

Ref: *https://365datascience.com/standardization/*

Normalization vs Standardization

- They both equals the **weights** (scale of magnitude) of features!
- **Normalization** is affected by **outliers** for its **denominator**!
- **Standardization** does not change the **shape** of the data.
- **If Distance-based: Standardization** (e.g, K-means, KNN, regressions, SVM, NN)
- **If Not Distance-based: Normalization** (e.g, Decision trees, Fisher LDA and Naive Bayes).

Ref: *https://sebastianraschka.com/Articles/2014_about_feature_scaling.html*

Normalization vs Standardization

- When in doubt, just **standardize** the data, it shouldn’t hurt!
- **Gaussian scaling** will be helpful if your data are roughly **normally distributed**.

Ref: *https://sebastianraschka.com/Articles/2014_about_feature_scaling.html*

End of Presentation

Presented by Berk Sudan