H-W 7

2)(a) $x_i \in R$

μ_i, σ_i denote mean and s.d till i^{th} sample resp

$$\mu_i = \frac{\sum_{k=1}^{i} x_k}{i} = \frac{\sum_{k=1}^{i-1} x_k + x_i}{i}$$

$$\boxed{\mu_i = \frac{(i-1)\mu_{i-1} + x_i}{i}} \quad \text{and} \quad \mu_1 = x_1$$

$$\sigma_i^2 = E(x^2)_i - E(x)_i^2 \qquad E(x^2)_i = \frac{(i-1)E(x^2)_{i-1} + x_i^2}{i}$$

$$= \frac{(i-1)\left[\sigma_{i-1}^2 + \mu_{i-1}^2\right] + x_i^2}{i}$$

$$\boxed{\sigma_i^2 = \frac{(i-1)\left[\sigma_{i-1}^2 + \mu_{i-1}^2\right] + x_i^2}{i} - \mu_i^2}$$

$$S.D, \ \sigma_i = \sqrt{\frac{(i-1)(\sigma_{i-1}^2 + \mu_{i-1}^2) + x_i^2}{i} - \mu_i^2}$$

(b) $x_i \in R^d$

Mean and Standard deviation will be calculated same as for $x \in R$ just μ will be d dimensional

$$\mu_i = \frac{(i-1)\mu_{i-1} + x_i}{i} \qquad [\mu_i \text{ will be } R^d]$$

$$\sigma_m^i = \sqrt{\frac{(i-1)\left((\sigma_m^{i-1})^2 + (\mu_m^{i-1})^2\right) + (x_m^i)^2}{i} - (\mu_m^{i-1})^2} \Rightarrow \sigma^i = \sqrt{\frac{(i-1)(\sigma^{i-1\,2} + \mu^{i-1\,2}) + (x^i)^2 - \mu^{i\,2}}{i}}$$

σ_m^i and μ_m^i denote the m^{th} component $(m \in [1, d])$ s.d and mean till i samples.

Let σ_{mn}^i denote covariance of m^{th} & n^{th} component till i samples $m \in [1, d]$ & $n \in [1, d]$

$$\sigma_{mn}^i = E(x_m x_n)^i - E(x_m)^i E(x_n)^i$$

Date: / /
Page:

$$E(x_m x_n)^i = \frac{(i-1)\,E(x_m x_n)^{i-1} + x_m^i x_n^i}{i}$$

$$= \frac{(i-1)\left[\sigma_{mn}^{i-1} + \mu_m^{i-1}\mu_n^{i-1}\right] + x_m^i x_n^i}{i}$$

$$\sigma_{mn}^i = \frac{(i-1)\left[\sigma_{mn}^{i-1} + \mu_m^{i-1}\mu_n^{i-1}\right] + x_m^i x_n^i}{i} - \mu_m^i \mu_n^i$$

(b) $x_i \in R$

let μ_i, σ_i denote Mean and s.d till i^{th} sample for the last M samples.

$$\mu_i = \mu_{i-1} + \left(\frac{x_i - x_{i-M}}{M}\right)$$

$$\sigma_i^2 = E(x^2)_i - E(x)_i^2$$

$$E(x^2)_i = E(x^2)_{i-1} + \frac{x_i^2 - x_{i-M}^2}{M} = \left(\sigma_{i-1}^2 + \mu_{i-1}^2\right) + \frac{x_i^2 - x_{i-M}^2}{M}$$

$$\sigma_i^2 = \left(\sigma_{i-1}^2 + \mu_{i-1}^2\right) + \frac{x_i^2 - x_{i-M}^2}{M} - \mu_i^2$$

$$\sigma_i = \sqrt{\sigma_{i-1}^2 + \left(\mu_{i-1}^2 - \mu_i^2\right) + \frac{x_i^2 - x_{i-M}^2}{M}}$$

$x_i^i \in R^d$

Let μ_m^i denote mean of m^{th} component till i samples

" σ_m^i " s.d " " " " "

" σ_{mn}^i " covariance b/w m^{th} & n^{th} component till i samples.

μ_m^i is calculated in the same way

$$\mu_m^i = \mu_m^{i-1} + \frac{x_m^i - x_m^{i-M}}{M} \Rightarrow \mu^i = \mu^{i-1} + \frac{x^i - x^{i-M}}{M}$$

$$\mu^i \in R^d$$

$$\sigma_m^i = \sqrt{(\sigma_m^{i-1})^2 + (\mu_m^{i-1})^2 - (\mu_m^i)^2 + \frac{(x_m^i)^2 - (x_m^{i-M})^2}{M}}$$

$$\sigma^i = \sqrt{(\sigma^{i-1})^2 + (\mu^{i-1})^2 - (\mu^i)^2 + \frac{(x^i)^2 - (x^{i-M})^2}{M}}$$

$$\sigma_{mn}^i = E(x_m x_n)^i - E(x_m)^i E(x_n)^i$$

$$E(x_m x_n)^i = E(x_m x_n)^{i-1} + \frac{x_m^i x_n^i - x_m^{i-M} x_n^{i-M}}{M}$$

$$= \sigma_{mn}^{i-1} + \mu_m^{i-1}\mu_n^{i-1} + \frac{x_m^i x_n^i - x_m^{i-M} x_n^{i-M}}{M}$$

$$\sigma_{mn}^i = \sigma_{mn}^{i-1} + \mu_m^{i-1}\mu_n^{i-1} + \frac{x_m^i x_n^i - x_m^{i-M} x_n^{i-M}}{M} - \mu_m^i \mu_n^i$$

~~In this way, in~~ ~~[illegible]~~ $\sigma^i = \sqrt{\frac{(i-1)((\sigma^{i-1})^2 + (\mu^{i-1})^2) + (x^i)^2}{i} - (\mu^i)^2}$

In this way we can calculate means, standard deviations & covariances from their corresponding previous value

Date : / /
Page :

Q3(a) For θ very small,

$$|y_i - w^T x| > \theta$$

The algorithm will be ineffective as very small θ, most samples in the dataset will be classified as outliers, even the samples which are not outliers and correct part of the dataset

For θ very large,
The algorithm will be ineffective as very large θ will even consider outliers as part of the dataset and won't rule them out of the dataset

(b) One way to compute θ can be to take θ such that for 95% of the data $|y_i - w^T x_i| > \theta$, here, we take the fact that outliers contribute to a small portion of the dataset. Here we assume 5% are outliers, this percentage can be varied a bit between 0 to 10%

Another way to compute θ can be to fix θ as mean of $|y_i - w^T x_i|$ or twice mean of $|y_i - w^T x_i|$ so that it can rule out outliers and also doesn't rule actual data as outliers.

Date : / /
Page :

Ans 1) Let $\underset{p\times 1}{y_i} = \underset{p\times d}{W^T} \underset{d\times 1}{x_i} + \epsilon_i$

Converting to vectors of each for all samples

$$\underset{N\times p}{Y} = \underset{N\times d}{X} \underset{d\times p}{W} + \underset{N\times 1}{E}$$

Loss function = $E^T E$

$$J(W) = [Y - XW]^T [Y - XW]$$

$$J(W) = Y^T Y - \underbrace{Y^T X W - W^T X^T Y}_{\text{Scalars} \Rightarrow A = A^T} + W^T X^T X W$$

$$\underbrace{J(W) = Y^T Y - 2 W^T X^T Y + W^T X^T X W}_{\text{All Scalars of dim } (1\times 1)}$$

$$\frac{\partial J(W)}{\partial W} = 0 \Rightarrow -2X^T Y + 2X^T X W = 0$$

$$\underset{d\times p}{W} = (\underset{d\times N}{X^T} \underset{N\times d}{X})^{-1} \underset{d\times N}{X^T} \underset{N\times p}{Y}$$

The closed form for W is same as for case of y scalar, it's just that W is now a matrix of $d\times p$ dimension, instead of a vector as earlier. We can get the same case of y_i scalar by substituting $p=1$, W is now $d\times p$ as y_i is $p\times 1$ instead of 1×1 as earlier.