Machine Learning - Homework 3

Parita Pooj (psp2133)

October 12, 2016

Problem 1

(a) **Centering**: No, the transformation does not affect the learning algorithm. Centering will basically make the mean $\hat{\boldsymbol{\mu}}$. This can be shown as below: Let $\hat{\boldsymbol{\mu}}'$ be the new mean parameter trained on the transformed data.
$\hat{\boldsymbol{\mu}}' = \frac{1}{n}\sum_{i=0}^{n}(\boldsymbol{x} - \hat{\boldsymbol{\mu}})$
$\hat{\boldsymbol{\mu}}' = \frac{1}{n}\sum_{i=0}^{n}\boldsymbol{x} - \dfrac{1}{n}\sum_{i=0}^{n}\hat{\boldsymbol{\mu}})$
$\hat{\boldsymbol{\mu}}' = \hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}$
$\hat{\boldsymbol{\mu}}' = 0$
Thus, centering essentially transforms the mean to zero, but the distribution still remains the same and hence, the classification won't be affected.

Standardization: No, standardization does not affect the learning algorithm. Standardization makes the standard deviation 1 for each feature, thus not affecting the classification, same as above.

(b) **Centering**: No, Centering preserves the order of the Euclidean distance between every pair of points. Hence, 1-NN classifier will not be affected. The preservation of the order of distances can be shown by considering three points $\boldsymbol{x_p}$, $\boldsymbol{x_q}$ and $\boldsymbol{x_r}$ such that:
$\sum_{i=1}^{n}(x_{p,i} - x_{q,i})^2 \leq \sum_{i=1}^{n}(x_{p,i} - x_{r,i})^2$
For transformed points, $\boldsymbol{x'_p}$, $\boldsymbol{x'_q}$ and $\boldsymbol{x'_r}$ we see that
$\sum_{i=1}^{n}(x'_{p,i} - x'_{q,i})^2 \leq \sum_{i=1}^{n}(x'_{p,i} - x'_{r,i})^2$
since,
$\sum_{i=1}^{n}(x_{p,i} - \hat{\mu}_i - x_{q,i} + \hat{\mu}_i)^2 \leq \sum_{i=1}^{n}(x_{p,i} - \hat{\mu}_i - x_{r,i} + \hat{\mu}_i)^2$

(c)

(d)

Problem 2

(a)

(b)

Problem 3

(a) The multivariate Gaussian distribution can be written as:

$$P_{\mu,\sigma^2} = \prod_{i=1}^{n} \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{\frac{-1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

where $\Sigma = \sigma^2 I$

$$\ln P_{\mu,\sigma^2} = \sum_{i=1}^{n} \frac{-1}{2}\ln(2\pi) - \frac{1}{2}\ln(|\Sigma|) - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$$

$$\ln P_{\mu,\sigma^2} = -\frac{1}{2}\sum_{i=1}^{n} \ln(2\pi) + \ln(|\Sigma|) + (\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$$

By matrix derivation rules[1],

$$\frac{\partial \ln P_{\mu,\sigma^2}}{\partial \Sigma} = -\frac{1}{2}\sum_{i=1}^{n} 0 + |\Sigma|^{-T} + (-\Sigma^{-T}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})\Sigma^{-T})$$

$$\frac{\partial \ln P_{\mu,\sigma^2}}{\partial \Sigma} = 0$$

$$-\frac{1}{2}\sum_{i=1}^{n}[|\Sigma|^{-T} - \Sigma^{-T}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-T}] = 0$$

Since, Σ is a diagonal matrix: $\Sigma^{-T} = \Sigma^{-1}$

$$\sum_{i=1}^{n}[|\Sigma|^{-1} - \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}] = 0$$

$$\sum_{i=1}^{n}|\Sigma|^{-1} = \sum_{i=1}^{n}[\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}]$$

$$\sum_{i=1}^{n} I = \sum_{i=1}^{n}[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}]$$

$$nI = \sum_{i=1}^{n}[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T]\Sigma^{-1}$$

$$n\Sigma = \sum_{i=1}^{n}[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T]$$

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n}[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T]$$

$$\sigma^2 I = \frac{1}{n}\sum_{i=1}^{n}[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T]$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}[\sum_{j=1}^{d}(\boldsymbol{x_j} - \boldsymbol{\mu_j})^2]$$

(b)

References

[1] `https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf`

[2] `http://cs229.stanford.edu/section/gaussians.pdf`