P1

(a) Centering: No.Because centering only changes the position of the distribution
Standardization: Yes.Because the variance of the distribution has changed.

(b) Centering: No.Because distance between the points have not changed.
Standardization: Yes.Because

$$D = \sqrt{\sum_{i=1}^{d}(u_i - v_i)^2}$$

Upon Standardization,

$$D = \frac{\sqrt{\sum_{i=1}^{d}(u_i - v_i)^2}}{\hat{\sigma}_i}$$

.

(c) Centering: No, because the relative position of the datas and the median of the datas have not changed.
Standardization: No, because the median of the points have not changed.

(d) Centering: No. Because as the data moves to the center, the classifier doesnt need to change
Standardization: No, the data will still be linearly seperable after standardization

P2

(a) The forth data representation I choose is bigram representation without any unigram features.

(b) PLA-Unigram:11.3575% / 11.4475%/ 10.7775%/ 11.8275%/ 11.3025%/ avg = 11.3425%
PLA-Bigram : 9.8225% / 9.885%/ 9.935%/ 10.2375%/ 9.345%/ avg = 9.845%
PLA-Chosen : 12.325% / 12.68%/ 11.9725%/ 12.045%/ 12.17%/ avg = 12.2385%
PLA-Tfidf : 11.4325% / 11.2575%/ 11.5125%/ 11.26%/ 11.4975%/ avg = 11.392%
Bayes-Unigram: 14.005% / 14.315%/ 14.205%/ 14.065%/ 14.0475%/ avg = 14.1275%

(c) Bigram representation

(d) Training ER = 8.42% , Test ER = 11.23%

P3

(a)

$$\mathcal{L} = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}$$

Get the derivative repect to σ.

$$0 = -\frac{n}{2}\cdot\frac{1}{\sigma^2} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4}$$

$$0 = -n\sigma^2 + \sum_{i=1}^{n}(x_i-\mu)^2$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$