Arda Türkoğlu-21601187

CS464 – Spring2020-HW1-Report

Q1.

Q1.1

$P(win) = 0.6$, $P(lose) = 0.4$ 8 matches left.

Suppose we had X wins and 8 − X losses, because we have win at least 1 to make turnover. (or X Loss, 8−x wins)

$$P(Turnover=1) = \sum_{x=1}^{7} P(win)^x P(Lose)^{8-x} + \sum_{x=1}^{7} P(Lose)^x P(win)^{8-x}$$

Same

$$= 2 \sum_{x=1}^{7} P(win)^x P(lose)^{8-x}.$$

$$= 2((0.6)^1(0.4)^7 + (0.6)^2(0.4)^6 + (0.6)^3(0.4)^5 + (0.6)^4(0.4)^4 + (0.6)^5(0.4)^3 + (0.6)^6(0.4)^2 + (0.6)^7(0.4)^1$$

$$= 0.632525$$

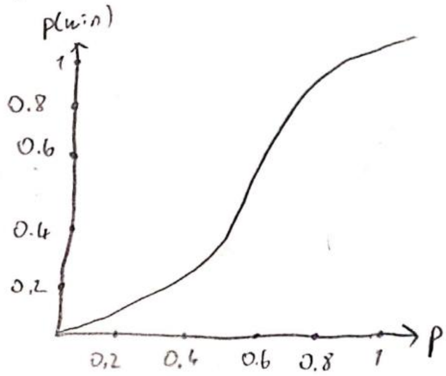
Q1.2

Our win condition is 2 successive points or point → miss → win or miss → score → win

$$\Rightarrow P(win) = \underbrace{P(score)}_{p} . \underbrace{P(score)}_{p} + \underbrace{P(score)}_{p}\underbrace{P(miss)}_{1-p}\underbrace{P(win)}_{?} + \underbrace{P(miss)}_{1-p}\underbrace{P(score)}_{p}\underbrace{P(win)}_{?}$$

$$= p^2 + 2p(1-p)P(win) = p^2 + (2p - p^2)P(win)$$

$$\boxed{P(win) = \frac{p^2}{2p^2 - 2p + 1}}$$



Plot calculated from wolfram,

Q3. For that question my implementation gives outputs of the forward selection and frequency selection. However, due to complexity issues in naïve_bayes(double for loops) run time is approximately ~20-25 minutes for each of them. I figured out using numpy arrays can get rid of the double for loops in naïve_bayes which increase efficiency. Unfortunately, I figured out that issue too late and could not spend time to fix that problem. Although it is working slow, it gives true outputs.

Q4.

Q4.1

w_1 is the unit vector and coveriance of the x is Σ.

We know that eigenvector with the largest eigenvalue has the maximum variance. $Var(z_1) = w_1^T \Sigma w_1$. where $cov(x) = \Sigma$

$z_1 = \langle w_1, x \rangle$

$$\max_{w_1} \frac{w_1^T \Sigma w}{w_1^T w} = \lambda_1 \quad , \quad \min_{w_1} \frac{w^T \Sigma w}{w_1^T w} = \lambda_{last}$$

$\Sigma x = \lambda x \Rightarrow \det(\Sigma - \lambda I) = 0$. λ_{max} gives max variance.

Thus, first component of PCA is eigenvector with λ_{max}.

$Z_1 = \alpha_{11} X_1 + \alpha_{21} X_2$ ———— $+ \alpha_{p1} X_p$, p is dimension

which is linear combination of features x. and has largest sample variance.

$\sum_{i=1}^{p} \alpha_{i1}^2 = 1$, $\alpha_1 = (\alpha_{11}, \alpha_{21} \ldots \alpha_{p1})^T$

We have to maximize α. α defines a direction in feature space where data changes the most.

Q4.2

Similary second component has form $z_2 = \alpha_{12} x_1 + \alpha_{22} x_2 + \alpha_{32} x_3 \ldots + \alpha_{p2} x_p$

According to the PCA, z_2 with direction α_2 must be orthogonal to the direction α_1 of z_1.

Referenced from book: "Introduction to Statiscal Learning" p.375-376