

Learning Coursework: Task 1 report

April 7, 2020

1.2 - Correlation matrix interpretation

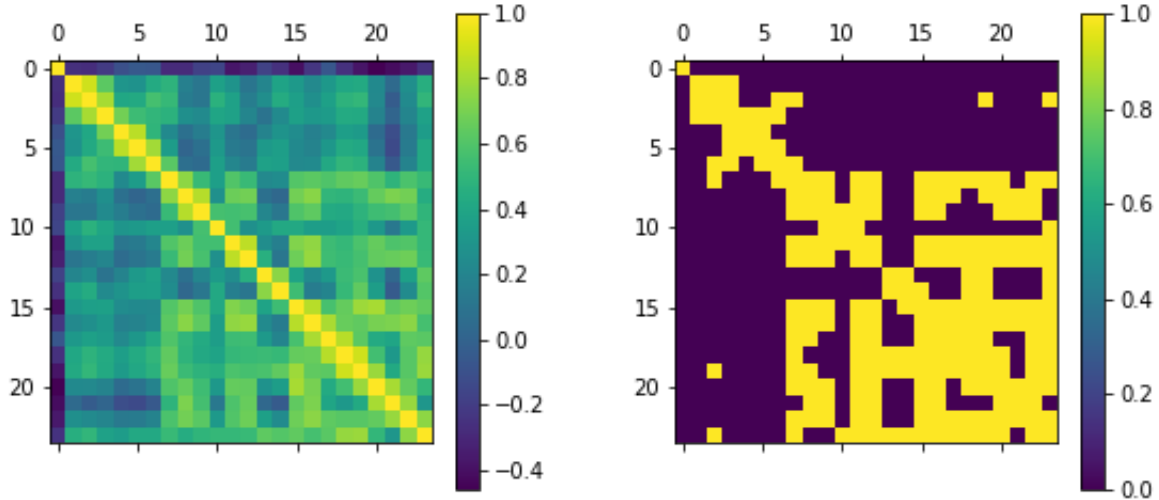


Figure 1: Covariance matrix heatmaps - the full one and one with an applied step function (1 if > 0.5 , -1 if < 0.5 , 0 otherwise)

Negative correlations

As we can see on Figure 1, the only feature that is negatively correlated with others is the feature with index 0 (dark blue bars on the upper and left edges of the matrix). These are weak negative correlations, as they are bounded below by about -0.4 .

Positive correlations

As we can see in the second plot on Figure 1, there are many features with high positive correlations (> 0.5 , seen in yellow). We can observe two clusters of features that are mostly correlated with each other - features in $[1, 6]$ and features in $[7, 23]$ (the rectangles that contain all the pairs of features within these intervals are mostly yellow).

1.3

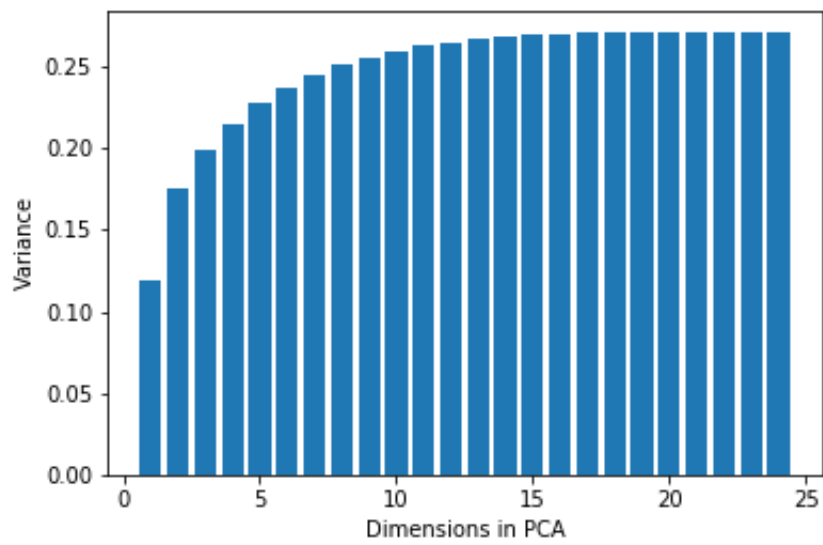


Figure 2: Task 1.3b - Cumulative variance graph

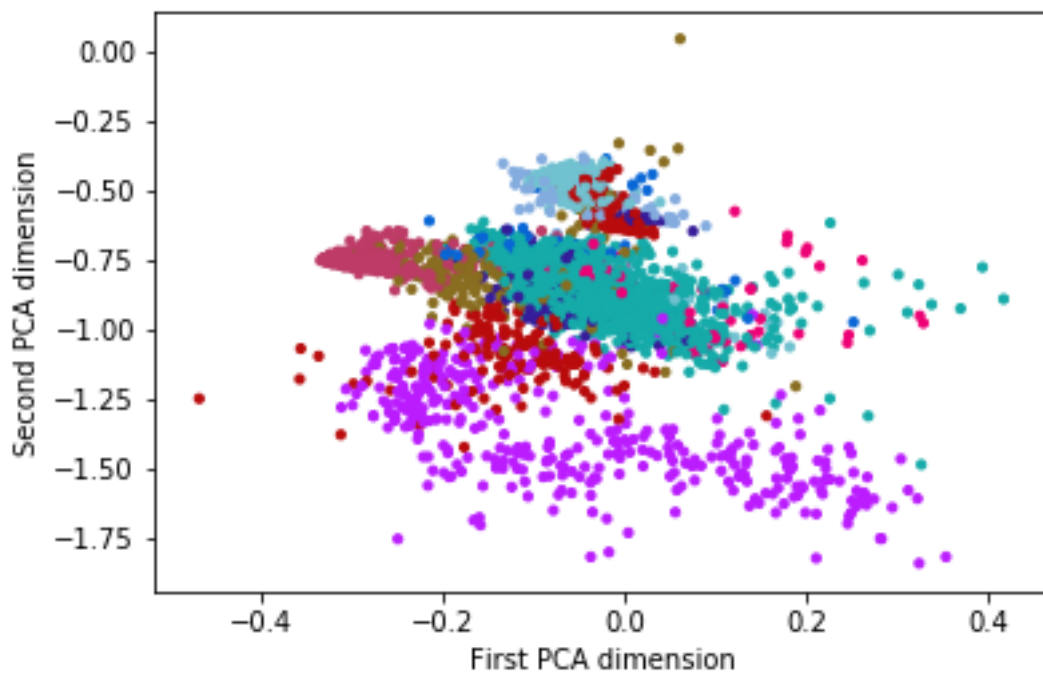


Figure 3: Task 1.3c - 2D-PCA plane

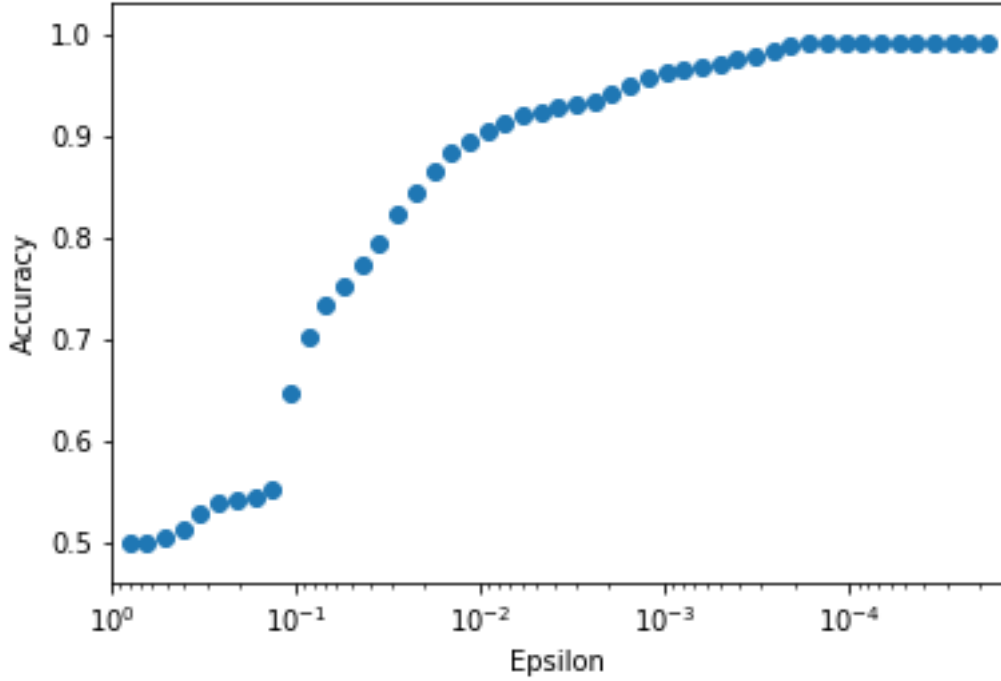


Figure 4: 1.5 - Impact of the epsilon on accuracy; logarithmically scaled x-axis

1.4b - Accuracy vs Covariance Matrix Type

CovKind	<i>Full</i>	<i>Diagonal</i>	<i>Shared</i>
Accuracy	90.101%	83.028%	88.270%

1.5 - Regularisation parameter analysis

Intuitively, we would expect the accuracy to drop when taking bigger epsilons, as this makes the approximation of the covariance matrix worse. Figure confirms that; accuracy increases when we take smaller epsilons. What is more, the plot becomes nearly flat for $x < 10^{-4}$, so there is no benefit from taking smaller epsilons than that.