<u>**Weekly report of lessons**</u>

Name: Ravi Pratap Singh
Roll No: 20CS60R60
The week : 20 oct ,21 oct , 22 oct

The topics covered: PCA properties , PCA applications , Linear discriminant analysis ,Better measure of separation ,Scatter matrix , Maximizing measure of separation ,Parametric method ,Maximum likelihood estimation,Bernoulli distribution , Gaussian distribution, bais ,mean and variance of estimator .

Summary topic wise :

- PCA diagonalizes the data covariance matrix as $\sum = CDC^T$, where D is diagonal matrix and C are unit column vectors of $\sum \rightarrow CC^T = C^T C = I$. Covariance among components is zero .
- Used for data compression (usually depends on data) . Helpful in decorelating components (like color image in RGB space , remote sensing images) . Factor analysis(useful for classification of uncorrelated factors).
- LDA is dimensionality reduction techniques reduce the number of dimensions (i.e. variables) in a dataset while retaining as much information as possible .It is used to capture maximum variance direction in the dataset.
- For better separation Normalize by a factor proportional to class variances . Scatter of data belonging to class C will be : $s^2 = \sum_{y \in c} (y - m_c)^2$, where *s* is class variance and m_c is mean . The measure of separation would be : $J(u) = \frac{D^2}{(s_1{}^2 + s_1{}^2)}$
- Scatter matrix : $S_C = \sum_{x \in C} (x - m_c) * (x - m_c)^T$. Within class scatter matrix : $S_w = S_1 + S_2$, where $S_1 = \sum_{x \in W_1} (u^T x - u^T m_1)\ (u^T x - u^T m_1)^T$
- To maximize $J(u)$, u should be such that $S_w{}^{-1} S_B u = \lambda u$,for any vector z , $S_B * z = k(m_1 - m_2)$, therefore $u = S_W{}^{-1}(m_1 - m_2)$.
- Parametric method includes a probability density function of known form and described by a set of parameters. It is used in bayesian inference.
- MLE of θ : $\theta^* = argmax\ l(\theta / x)$. Log likelihood $log(l(\theta/x)) = L(\theta/x) = \sum_{t=1}^{N} logP(x^t/\theta)$
- Bernoulli distribution has Likelihood $L(p/X) = log\ p * (\prod_{t=1}^{N} x^t) + log(1-p) * (N - \sum_{t=1}^{N} x^t)$. Maximizing this function we get $\overline{p} = \frac{\sum_{t=1}^{N} x^t}{N}$
- Gaussian density , For a gaussian distribution $p(x)$ in range $-\alpha\ to\ +\alpha$ with expectation $E(X) = \mu$, MLE will be : $\sigma^2 = s^2 = \frac{\sum_{t=1}^{N}(x^t - m)^2}{N}$.
- The bias of an estimator is the difference between this estimator's expected value and the true value of the parameter being estimated i.e $b_\theta(d) = E(d(X) - \theta)^2)$ $\leftarrow$ This is also the MSE of estimator.

  Mean : $m = \frac{1}{N} \sum_{t=1}^{N} x^t$; expectation of mean : $E(m) = \frac{1}{N} \sum_{t=1}^{N} E(x^t) = \frac{N\mu}{N} = N$; Variance of mean: $var(m) = \frac{\sigma^2}{N}$

  Variance : $var(x) = E(X^2) - (E(X))^2$ .We consider sample variance as estimate of variance of distribution so , $E(s^2) = \frac{N-1}{N} \sigma^2$ $\leftarrow$ This is asymptotically unbiased because N is very large in real examples .

Any novel idea of yours out of the lessons : PCA could be used with other advanced compression techniques and transformations to give better performance .