EE 657 - Pattern Recognition and Machine Learning

Problem Set

---

1. Consider a dataset in two dimensions where the data lies on the circumference of a circle of unit radius. What would be the effect of using PCA on this dataset, in which we attempt to reduce the dimensionality to 1? Suggest an alternative one dimensional representation of the data.

2. Let S be the covariance matrix of the data. The Mahalanobis distance between x^a and x^b is defined as
$$(x^a - x^b)^T S^{-1}(x^a - x^b)$$
Explain how to approximate this distance using M-dimensional PCA approximations.

3. Consider the following 3-dimensional datapoints:
$$(1:3;1:6;2:8)(4:3;1:4;5:8)(0:6;3:7;0:7)(0:4;3:2;5:8)(3:3;0:4;4:3)(0:4;3:1;0:9)$$
Perform Principal Components Analysis by:
    1. Calculating the mean, c, of the data.
    2. Calculating the covariance matrix $S = \frac{1}{6}\sum_{n=1}^{6} x^n(x^n)^T \; - \; cc^T$ of the data.
    3. Finding the eigenvalues and eigenvectors e^i of the covariance matrix.

4. Consider the distribution dened on real variables x; y:
$$p(x,y) \propto (x^2+y^2)^2 e^{-x^2-y^2}, \quad dom(x) = dom(y) = \{-\infty \dots \infty\}$$
Show that $E[x] = E[y] = 0$. Furthermore show that x and y are uncorrelated, $E[xy] = E[x]E[y]$. Whilst x and y are uncorrelated, show that they are nevertheless dependent.

5. For a variable x with $dom(x) = \{0, 1\}$, and $p(x = 1) = \theta$, show that in n independent draws $x_1, \dots, x_n$ from this distribution, the probability of observing k states 1 is the Binomial distribution
$$\binom{n}{k}\theta^k(1-\theta)^{n-k}$$

6. The normalisation constant of a Gaussian distribution is related to the integral
$$I = \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx$$
By considering
$$I^2 = \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy = \int_{-infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2+y^2} dxdy$$
and transforming to polar coordinates, show that
    1. $I = \sqrt{2\pi}$.
    2. $\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = \sqrt{2\pi\sigma^2}$

7. For a univariate Gaussian distribution, show that
    1. $\mu = E[x]_{\mathcal{N}(x|\mu,\sigma^2)}$.
    2. $\sigma^2 = E[(x-\mu)^2]_{\mathcal{N}(x|\mu,\sigma^2)}$

8. Using
$$E[x^T Ax] = E[trace(Axx^T)] = trace(A\ E[xx^T])$$
derive $E[x^T Ax]_{\mathcal{N}(x|\mu,\Sigma)} = \mu^T A\mu + trace(A\Sigma)$.

9. Consider

$$I = \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} dx$$

By using the transformation

$$z = \Sigma^{-\frac{1}{2}}(x - \mu)$$

show that $I = \sqrt{\det(2\pi\Sigma)}$.

10. For a Gaussian mixture model

$$p(x) = \sum_i p_i \mathcal{N}(x|\mu_i, \Sigma_i), \quad p_i > 0, \sum_i p_i = 1$$

show that $p(x)$ has mean

$$E[x] = \sum_i p_i \mu_i$$

and covariance

$$\sum_i p_i(\Sigma_i + \mu_i m u_i^T) - \sum_i p_i \mu_i \sum_j p_j \mu_j^T$$

11. Consider data x^n, $n = 1, \ldots, N$. Show that for a Gaussian distribution, the Maximum Likelihood estimator of the mean is $\hat{m} = \frac{1}{N}\sum_{n=1}^{N} x^n$ and variance is $\hat{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N}(x^n - \hat{m})^2$.