Assignment 3

1. Naive Bayes Classification: Calculate the likelihood and prior for training data and use it to classify a given test data using Baye's Theorem.
   Observation Accuracies:
   78.505086245 %
   77.620521893 %
   78.062804069 %
   77.7532065458 %
   77.0455550641 %
   77.5762936754 %
   77.6647501106 %
   77.0455550641 %
   77.3109243697 %
   77.4878372402 %

   Mean Accuracy: 77.607253428 %
   Standard Deviation: 0.133133547

   Though the chances of a tie is very less since log of probabilities gives a real value, but in case of a tie, the "yes" class is chosen always
   For now, the missing entries have been ignored in the posterior calculations

2. In the end

3. <u>PCA:</u>
   The mean vector and covariance vector is calculated first. Then, from the covariance matrix, eigenvectors and eigenvalues are obtained. The eigenvectors corresponding to k largest eigenvalues is taken to form a matrix W. W^TX gives us the new feature set where X is the old feature set.
   Accuracies:
   A. K=10 → 68%
   B. K=100 → 56%
   C. K=1000 → 55%

   <u>LDA:</u>
   Calculate the mean for each class and then calculate within class and between class scatter matrix. From these two, obtain a matrix ($S_w^{-1}S_b$) and calculate its eigenvalues. Take the largest eigenvalue and this is the required weight vector for LDA.
   Accuracy: 83%
   <u>Gaussian Bayesian Classifier:</u>
   In this classifier, mean vector and covariance matrix is computed for each class and for a particular test data, multivariate gaussian formula is used to calculate the likelihood for the test data. The data is classified by calculating posterior from likelihood and prior.

unknown Parametric form - We assume that it has known parametric form θ. ~~Any info~~ We know $p(x|\theta)$. Any information we might about θ is known prior density $p(\theta)$, obⁿ of samples converts it to a posterior $p(\theta|D)$ which we hope sharply peaks at the value of θ.

$$p(x, D) = \int p(x, \theta | D)\, d\theta$$
$$= \int p(x|\theta)\, p(\theta|D)\, d\theta$$

BPE : Gaussian case
$p(x, \mu) \sim N(\mu, \Sigma)$
~~Consider the case where μ is the only unknown parameter.~~
Univariate Case : $p(\mu|D)$
μ only unknown
$$p(x|\mu) \sim N(\mu, \sigma^2)$$

We assume that prior knowledge we might have about μ can be expressed by $p(\mu)$

$$p(\mu|D) = \cancel{p(D) \times p} \quad \frac{p(D|\mu) \times p(\mu)}{\int p(D|\mu) \times p(\mu)\, d\mu}$$
$$= \alpha \prod_{k=1}^{n} p(x_k|\mu)\, p(\mu)$$

α - factor that depends on D but not on μ

We assume $p(\mu) \sim N(\mu_0, \sigma_0^2)$ where μ_0, σ_0 are known

Scanned by CamScanner

$$p(\mu|D) = \alpha \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_k-\mu}{\sigma}\right)^2} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2}$$

$$\cancel{= \alpha' e^{-\frac{1}{2}\sum\left(\frac{x_k-\mu}{\sigma}\right)^2} +}$$

$$\cancel{= \alpha' e^{-\frac{1}{2}\left(\sum\left(\frac{x_k-\mu}{\sigma}\right)^2 + \left(\frac{\mu-\mu_0}{\sigma_0}\right)^2\right)}}$$

$$\cancel{= \alpha'' e\left[-\frac{1}{2}\left(\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)\mu^2 - 2\left(\frac{1}{\sigma^2}\sum x_k\right.\right.\right.}$$

$$= \alpha' e^{-\frac{1}{2}\left[\frac{\sum x_k^2 + \sum \mu^2 - \sum 2x_k\mu}{\sigma^2} + \frac{\mu^2+\mu_0^2-2\mu\mu_0}{\sigma_0^2}\right]}$$

$$p(\mu|D) = \alpha'' e^{-\frac{1}{2}\left[\mu^2\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right) - 2\mu\left(\frac{\sum x_k}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right)\right]}$$

~~Absorbing~~ Remaining terms indⁿ of μ

$$p(\mu, D) = N(\mu_n, \sigma_n^2)$$

$$\frac{1}{\sigma_n^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \Rightarrow \boxed{\sigma_n^2 = \frac{\sigma_0^2\sigma^2}{\sigma^2 n + \sigma^2}}$$

$$\frac{\mu_n}{\sigma_n^2} = \frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}$$

$$\bar{\mu} = \frac{n}{n\sigma_0^2+\sigma^2}\left[n\bar{x}\sigma_0^2 + \mu_0\sigma^2\right]$$

$$\mu_n = \left[\frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2}\right]\bar{x} + \left[\frac{\sigma^2}{n\sigma_0^2+\sigma^2}\right]\mu_0$$

As $\sigma_n^2 \to 0$, $n \to \infty$, each addⁿ observation decreases our uncertainty about μ. As n increases $p(\mu|D)$ becomes more & more sharply peaked, becoming delta function. This is also known as Bayesian

Scanned by CamScanner

In general μ_n is a linear comb^n of μ_i & μ_0 & lying b/w μ_i & μ_0. If $\sigma_0 = 0$ $\mu = \mu_n$, not the obs^n can change our opinion.

If $\sigma_0 >> \sigma$, $\mu_n = \mu_i$

$\frac{\sigma}{\sigma_0^2} \rightarrow$ degenerates

If not infinite after enough samples are taken exact values assumed for μ_0 & σ_0^2 unimportant μ_n converges to sample mean.

The Univariate Case: $p(x/D)$

$$p(x/D) = \int p(x/\mu)\, p(\mu/D)\, d\mu$$

$$= \int \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left[\frac{(x-\mu)}{\sigma}\right]^2} \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{1}{2}\left[\frac{\mu-\mu_n}{\sigma_n}\right]^2} d\mu$$

$$= \frac{1}{2\pi\sigma\sigma_n} e^{-\frac{1}{2}\left[\frac{x-\mu_n}{\sigma^2+\sigma_n^2}\right]^2} f(\sigma, \sigma_n)$$

$$f(\sigma, \sigma_n) = \int e^{-\frac{1}{2}\left[\frac{\sigma^2+\sigma_n^2}{\sigma^2\sigma_n^2}\right]\left(\mu - \frac{\sigma_n^2 x + \sigma^2\mu_n}{\sigma^2+\sigma_n^2}\right)^2}$$

Multivariate Case

$p(x/\mu) \sim N(\mu, \Sigma)$ & $p(\mu) \sim N(\mu_0, \Sigma_0)$

$$p(\mu/D) = \frac{p(D/\mu) \times p(\mu)}{\int p(D/\mu)\, p(\mu)\, d\mu}$$

$$= \prod_k p(x_k/\mu) \times p(\mu)$$

Scanned by CamScanner

$$p(\mu/D) = \alpha \prod p(x_k/\mu) \times p(\mu)$$

$$= \cancel{\alpha \prod \frac{1}{2\pi \Sigma}}$$

$$= \frac{1}{(2\pi)^{d/2} |\Sigma|^{\frac{1}{2}}} \prod e^{-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)} \times \frac{1}{(2\pi)^{d/2} |\Sigma_0|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mu-\mu_0)^t \Sigma_0^{-1}(\mu-\mu_0)}$$

$$p(\mu/D) = \alpha\, e^{-\frac{1}{2}\left[\sum (x-\mu)^t \Sigma^{-1}(x-\mu) + (\mu-\mu_0)^t \Sigma_0^{-1}(\mu-\mu_0)\right]}$$

$$= \alpha\, e^{-\frac{1}{2}\left[\sum (x^t\Sigma^{-1} - \mu^t\Sigma^{-1})(x-\mu) + (\mu^t\Sigma_0^{-1} - \mu_0^t\Sigma_0^{-1})(\mu-\mu_0)\right]}$$

$$= \cancel{\alpha e^{-\frac{1}{2}}\left[\sum (x^t\Sigma^{-1}x - x^t\Sigma^{-1}\mu - \mu^t\Sigma^{-1}x + \mu^t\Sigma^{-1}\mu) + \mu^t\Sigma_0^{-1}\mu - \right.}$$

$$\cancel{\alpha e^{-\frac{1}{2}}\left(\sum x^t\Sigma^{-1}x - \sum x^t\Sigma^{-1}\mu - \right.}$$

$$\alpha e^{-\frac{1}{2}\left(\sum x^t\Sigma^{-1}\mu - \sum \mu^t\Sigma^{-1}x + \sum \mu^t\Sigma^{-1}\mu + \mu^t\Sigma_0^{-1}\mu + \mu^t\Sigma_0^{-1}\mu_0 - \mu_0^t\Sigma_0^{-1}\mu\right)}$$

$$= \cancel{\alpha' e^{-\frac{1}{2}}\left(\mu^t(n\Sigma^{-1} + \Sigma_0^{-1})\mu - \mu^t(\Sigma_{2k})\right.}$$

$$- \mu^t\left(\Sigma^{-1}\sum x_k + \right.$$

$$= \alpha' e^{-\frac{1}{2}\left[\sum (\mu^t\Sigma^{-1}\mu - \mu^t\Sigma^{-1}x - x^t\Sigma^{-1}\mu) + \mu^t\Sigma_0^{-1}\mu - \mu^t\Sigma_0^{-1}\mu_0 - \mu^t\Sigma_0^{-1}\mu_0\right]}$$

Note $(\mu^t\Sigma^{-1}x)^t = x^t\Sigma^{-1}\mu$

$$= \alpha' e^{-\frac{1}{2}\left[\mu^t(n\Sigma^{-1} + \Sigma_0^{-1})\mu - 2\mu^t\left[\Sigma^{-1}\Sigma_0 x_k + \Sigma_0^{-1}\mu_0\right]\right]}$$

Scanned by CamScanner

$$= x'_{exp} -\frac{1}{2}\Big(\mu^t (n\Sigma^{-1} + \Sigma_0^{-1})\mu - 2\mu^t(\Sigma^{-1}\Sigma_{x_i} + \Sigma_0^{-1}$$

$$\cancel{e^{-\frac{1}{2}(x-\mu_n)^t \Sigma_n^{-1}(x-\mu_n)}}$$

$$e^{-\frac{1}{2}(\mu-\mu_n)^t \Sigma_n^{-1}(\mu-\mu_n)}$$

$$\Rightarrow \Sigma_n^{-1} = n\Sigma^{-1} + \Sigma_0^{-1}$$

$$\Sigma_n^{-1}\mu_n = n\Sigma^{-1}\bar{\mu} + \Sigma_0^{-1}\mu_0$$

$$(A^{-1} + B^{-1})^{-1} = A(A+B)^{-1}B$$

$$\mu_n = \Sigma_0\left(\Sigma_0 + \frac{1}{n}\Sigma\right)^{-1}\mu_n^2 + \frac{1}{n}\Sigma\left(\Sigma_0 + \frac{1}{n}\Sigma\right.$$

Scanned by CamScanner