Anomaly detection using multivariate gaussian distribution

Parameters - μ, Σ ; $\mu \in R^n$ $\Sigma \in R^{n \times n}$

$$p(x; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

$$\mu = \frac{1}{m} \sum_{i=1}^{m} x^{(i)}$$

$$\Sigma = \frac{1}{m} \sum_{i=1}^{m} (x^{(i)}-\mu)(x^{(i)}-\mu)^T$$

RECOMMENDER SYSTEMS

recommends projects or items

n_u = no. of users

n_m = no. of movies

$r(i,j) = 1$ if j has rated movie i

$y(i,j)$ = rating given by user j to movie i (defined iff $r(i,j) = 1$)

" Recommender ~~system~~ should predict the movie ratings that are not ~~filled~~ given by users & also predict what else can the user watch.

Content-based recommender system.

Movie	Alice	Bob	Carol	Dave	x_1 (romance)	x_2 (action)
Romcom 1	5	5	0	0	0.9	0
Romcom 2	5	?	?	0	1.0	0.01
Act 3	?	4	0	?	0.99	0
Act 4	0	0	5	4	0.1	1.0
Act 5	0	0	5	?	0	0.9

→ for each user j, learn $\theta^{(j)} \in \mathbb{R}^3$ $(\mathbb{R}^{n+1})$. predict user j's rating movie i with $(\theta^{(j)})^T x^{(i)}$

Problem formulation

$r(i,j) = 1$ if user j has rated movie i (0 otherwise)
$y(i,j)$ = rating by user j on movie i (if defined)

$\theta^{(j)}$ = parameter vector for user j
$x^{(i)}$ = feature vector for movie i
For user j, movie i, predicted rating = $(\theta^{(j)})^T x^{(i)}$

$m^{(j)}$ = no. of movies rated by user j

To learn $\theta^{(j)}$: (its linear regression)

$$= \min_{\theta^{(j)}} \frac{1}{2\cancel{m^{(j)}}} \sum_{i: r(i,j)=1} \left((\theta^{(j)})^T (x^{(i)}) - y^{(i,j)} \right)^2 + \frac{\lambda}{2\cancel{m^{(j)}}} \sum_{k=1}^{n} (\theta_k$$

$$\therefore = \min_{\theta^{(j)}} \frac{1}{2} \sum_{i: r(i,j)=1} \left((\theta^{(j)})^T x^{(i)} - y^{(i,j)} \right)^2 + \frac{\lambda}{2} \sum_{k=1}^{n} (\theta_k^{(j)})^2$$

That's for only one θ.

For $\theta^{(1)}, \theta^{(2)}, \ldots \theta^{(n_u)}$:

$$\min_{\theta^{(1)} \ldots \theta^{(n_u)}} \frac{1}{2} \underbrace{\sum_{j=1}^{n_u}}_{\text{for all user } j} \underbrace{\sum_{i: r(i,j)=1}}_{\text{sum over } i \text{ movies that the user has watched}} \left((\theta^{(j)})^T x^{(i)} - y^{(i,j)} \right)^2 + \frac{\lambda}{2} \sum_{j=1}^{n_u} \sum_{k=1}^{n} (\theta_k^{(j)})^2$$

$J(\theta^{(1)}, \ldots \theta^{(n_u)})$

<u>Gradient descent update</u>:

$$\theta_k^{(j)} := \theta_k^{(j)} - \alpha \sum_{i: r(i,j)=1} \left((\theta^{(j)})^T x^{(i)} - y^{(i,j)} \right) x_k^{(i)} \quad (\text{for } k=0)$$

$$\theta_k^{(j)} := \theta_k^{(j)} - \alpha \left(\sum_{i: r(i,j)=1} \left((\theta^{(j)})^T x^{(i)} - y^{(i,j)} \right) x_k^{(i)} + \lambda \theta_k^{(j)} \right)$$

$\ldots$ for $(k \neq 0)$