

Matrix Factorization

交互最小二乗法によるパラメータ最適化

$$R \approx P \times Q = \hat{R}$$

R は U 人のユーザ、 D 個のアイテムに対するratingのマトリクス \hat{R} はratingの予測値のマトリクス
ここで、この R を最もよく表す $P(|U| \times K$ 行列)と $Q(|D| \times K$ 行列)を求めたい。 K は近似する際の行列のランク

ユーザ i のアイテム j に対する予測値は以下で表される

$$\hat{r}_{ij} = p_i^T q_j = \sum_{k=1}^k p_{ik} q_{kj}$$

ユーザごとのバイアス値 u_i とアイテムごとのバイアス値 v_j を足し合わせることで、より精度の良い結果が得られることがある。

$$\hat{r}_{ij} = \sum_{k=1}^k p_{ik} q_{kj} + u_i + v_j$$

一旦簡略化のために、上記の正則化項は忘れる。 このときの P と Q を推定する際に最適化の対象となる目的関数は、元のrating値 r_{ij} と予測値 \hat{r}_{ij} の二乗和 E を用いる。

$$E = \sum (r_{ij} - \hat{r}_{ij})^2 = \sum (r_{ij} - p_i^T q_j)^2 = \sum (r_{ij} - \sum_{k=1}^k p_{ik} q_{kj})^2$$

交互最小二乗法を用いて学習を行う。 交互最小二乗法においては、求める P と Q の片方を固定したときに誤差 E を最小化するようにもう一方を推定することを繰り返し行う。

P を固定して Q を推定する。 各要素 q_{lj} ($l = 1, 2, \dots, K$)の偏微分が0になるように各要素を推定すればよい。

$$\frac{\partial E}{\partial q_{lj}} = \sum \left[\frac{\partial e_{ij}}{\partial q_{lj}} \right] = 2 \sum \left[p_{il} (r_{ij} - \sum_{k=1}^K p_{ik} q_{kj}) \right] = 0$$

また、次のステップでは、逆に Q を固定して P を推定する。

$$\frac{\partial E}{\partial p_{il}} = 2 \sum \left[q_{lj} (r_{ij} - \sum_{k=1}^K p_{ik} q_{kj}) \right] = 0$$

Q を固定する場合で考える。 \mathbf{q}_j の推定は、以下の様にかける。

$$\begin{bmatrix} \sum p_{i1}^2 & \cdots & \sum p_{i1} p_{iK} \\ \vdots & \ddots & \vdots \\ \sum p_{iK} p_{i1} & \cdots & \sum p_{iK}^2 \end{bmatrix} \begin{bmatrix} q_{j1} \\ \vdots \\ q_{jK} \end{bmatrix} = \begin{bmatrix} \sum s_{ij} p_{i1} \\ \vdots \\ \sum s_{ij} p_{iK} \end{bmatrix}$$

上記から、アイテム l に対する推定は、他のアイテムに対する推定と独立して行う事ができることがわかる。

過学習を避けるための手段として、正則化項を入れる場合、推定式は以下となる。

$$E = \sum (r_{ij} - p_i^T q_j^T)^2 + \lambda (\|\mathbf{q}_j\|^2 + \|\mathbf{p}_i\|^2)$$

$$\frac{1}{2} \frac{\partial E_{ij}}{\partial q_{lj}} = p_{il} \left(r_{ij} - \sum_{k=1}^K p_{ik} q_{kj} \right) + \lambda q_{lj}$$

$r_{ij} - \sum_{k=1}^K p_{ik} q_{kj}$ は固定値なので、 γ とおくと、

$$\frac{\partial E_{ij}}{\partial q_{lj}} = \gamma p_{il} + \lambda q_{lj}$$