Некоррелированные ответы

$$\sum_{i=1}^{n}(y_i - x_i B)(y_i - x_i B)^T \to \min_B,$$

где $y_i = (y_i^1, \dots, y_i^K)$ — строка ответов для i-го объекта, $x_i = (x_i^1, \dots, x_i^D)$ — признаковое описание i-го объекта, B — настраиваемая матрица весов.

Легко доказать, что в этом случае ответ не изменится и будет K независимых линейных моделей

$$\hat{B} = (X^T X)^{-1} X^T Y$$

Коррелированные ответы

$$\sum_{i=1}^{n}(y_i - x_iB)\Sigma^{-1}(y_i - x_iB)^T \to \min_{B},$$

где $y_i = (y_i^1, \ldots, y_i^K)$ — строка ответов для i-го объекта, $x_i = (x_i^1, \ldots, x_i^D)$ — признаковое описание i-го объекта, B — настраиваемая матрица весов, а Σ — матрица ковариации ответов.

Магия

$$\sum_{i=1}^{n}(y_i - Bx_i)\Sigma^{-1}(y_i - Bx_i)^T =$$

$$= \operatorname{tr}\left((Y - XB)\Sigma^{-1}(Y - XB)^T\right) =$$

Заметим, что $\exists L \colon \Sigma^{-1} = LL^T$

$$= \operatorname{tr}\left(((Y - XB)L)((Y - XB)L)^T\right) =$$

Введём новые переменные $\overline{Y} = YL$ и $\overline{B} = BL$

$$= \operatorname{tr}\left((\overline{Y} - X\overline{B})(\overline{Y} - X\overline{B})^T\right)$$

Мы свели задачу к предыдущему случаю

Магия

$$\operatorname{tr}\left((\overline{Y} - X\overline{B})(\overline{Y} - X\overline{B})^T\right) \to \min$$

$$\hat{\overline{B}} = (X^T X)^{-1} X^T \overline{Y}$$

$$\hat{B} L = (X^T X)^{-1} X^T Y L$$

$$\hat{B} = (X^T X)^{-1} X^T Y$$

Получилось точно такое же решение!