

# Simple Recommender System

Tony Kovanen

27.1.2012

## Sisältö

**Problem Description** The aim is a simple recommender system that recommends new movies to the user knowing only a list of movies and some features associated with the films (a coefficient between 0 and 1 that tells how strongly it can be assigned to some category). The user can rate movies on a scale from 0 to 5 stars. The program uses linear regression to fit a linear function  $h_{\theta}(x)$  to the data and predict ratings for films that have not been watched and rated yet.

**Representing the Data** Movies known by the system are put into two lists. Movies rated by the user (denoted as  $r$ ), and movies not rated by the user ( $r'$ ). Rated films can also be represented as a matrix  $X$ , where each row of the matrix represents a certain movie.

**Predicting Ratings** Our hypothesis function  $h_{\theta}(x)$  is how we predict a rating for an unrated movie.  $h_{\theta}(x) = \sum_{j=0}^m \theta_j x_j$  which is the same as  $h_{\theta}(x) = \theta^t x$  represented with vectors. Next we'll look at how we get the parameters  $\theta$ .

**Finding the parameters  $\theta$**  The parameters  $\theta$  can be found with various different methods. We want to use the normal equation method.  $\theta = (X^t X)^{-1} X^t y$

**The Program as a Whole** The program as a whole works in iterations of the same functions. First with some method (not decided yet) we will choose films to the new user in a sense that will give as a good idea of what he/she likes (controversial with each other). Then we will minimize the parameters  $\theta$  using the normal equation method. Then we will predict ratings for all unrated movies using the hypothesis function  $h_{\theta}(x)$ . We will then pick  $k$  (at the moment an undecided value).