Movie Recommendations Using Low-dimensional Codes and User Specified Feature Relevance

Chris Curro, David Katz, Harrison Zhao

May 2015

Goal

- Rich low-dimensional latent space
- High quality movie recomendations based on nearest neighbors
- User input on the reocomendation

Dataset

Tag Genome dataset provied by grouplens

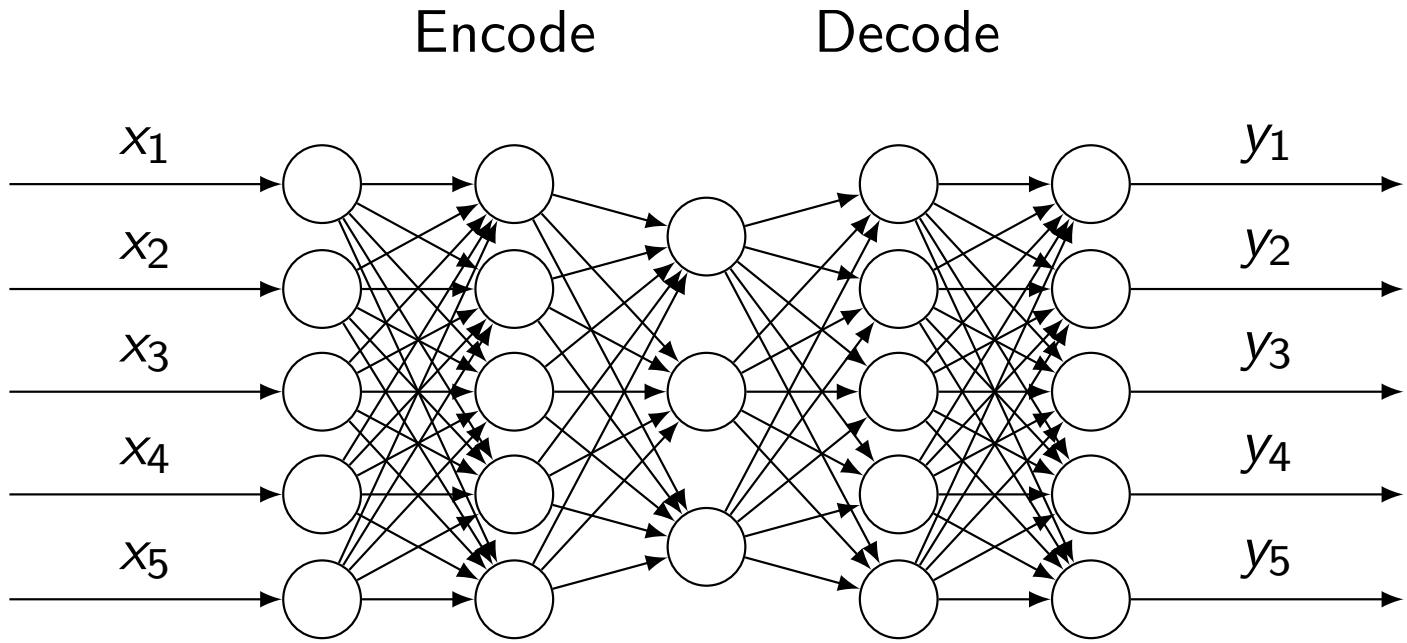
- 10,000 movies
- 1,000 tags
- relevance of each tag to each movie

Approach

- Autoencoders for nonlinear PCA
- Nearest neighbors with weighted features

Autoencoders

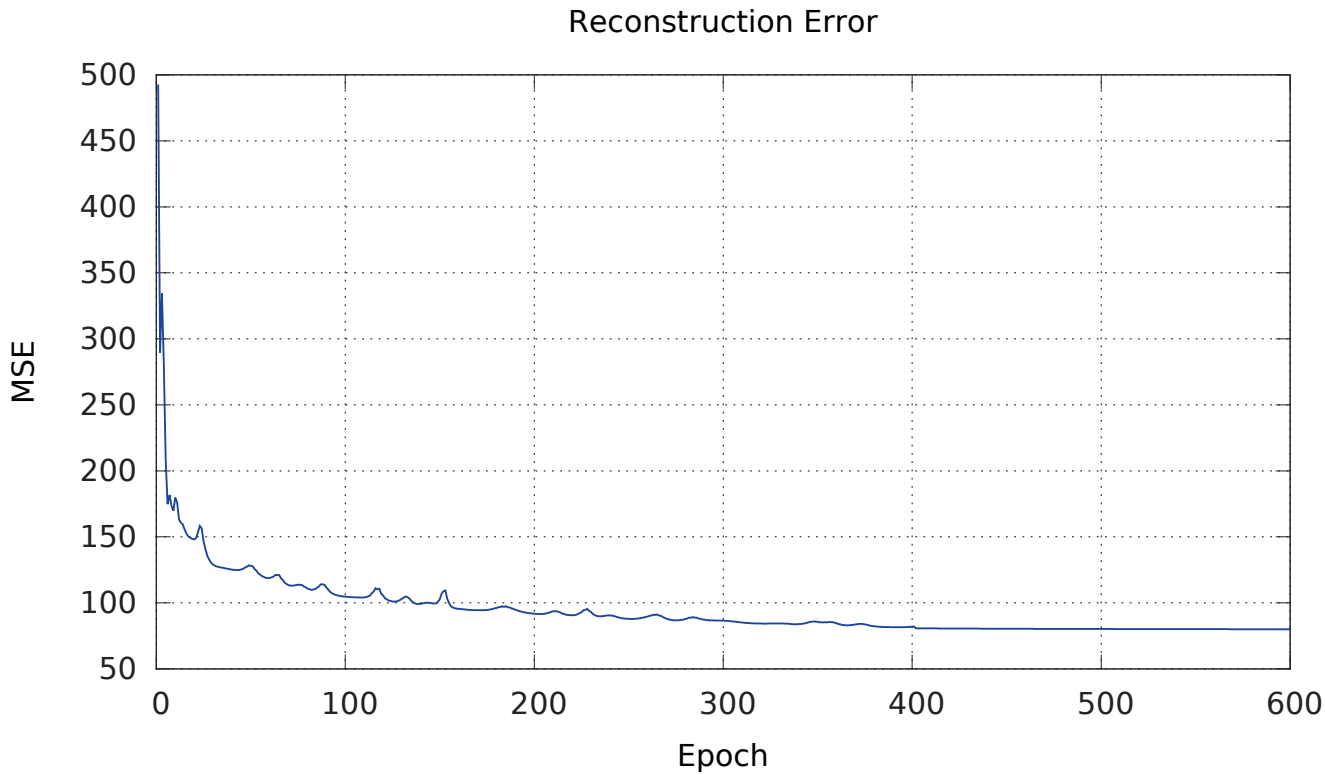
$$\mathbf{y} = f\left(W\mathbf{x} + \mathbf{b}\right) \tag{1}$$



Autoencoders

- Map 1128 feature space onto a 10 dimensional space
- Single hidden layer encoder and decoder
- Hyperbolic tangent neurons

Autoencoders



Nearest Neighbors

- Select the top movies to recommend via nearest neighbors
- Use modified distance metric which lets users specify feature relevance

$$distance(x, y, v) = \sqrt{\sum_{i=1}^{D}((x_i - y_i) \times v_i \times D)^2} \qquad (2)$$

Approximate Nearest Neighbors

- A linear search is expensive on a large dataset with many users
- No algorithm can find exact nearest neighbors in sub-linear time for high dimensional vector spaces
- ANN indexes can return a high quality results from only examining a small percentage of the search space
- Problem: Conventional ANN indexes are not adaptable to query specified dimension relevance
- Solution: An Efficient Index for Computation of Approximate Nearest Neighbors with Query Specified Dimension Relevance Weights

Performance

Difficult to evaluate without explicit user feedback

Demo

Movie Recommendations Using Low-dimensional Codes and User Specified Feature Relevance

Chris Curro, David Katz, Harrison Zhao

May 2015