PCA

1. Assume we are given three data points in $\mathbb{R}^3$: $(1,1),(2,2),(3,3)$.

   (a) Compute the covariance matrix. Recall the following formula:

   $$\Sigma[t,s] = \frac{1}{n-1}\sum_{i=1}^{n} x_{c,i}[t]x_{c_i}[s]$$

   Solution.

   $$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

   (b) The associated eigenvalues and eigenvectors of the previous matrix are

   $$\lambda_1 = 2, \lambda_2 = 0, v_1 = (1,1), v_2 = (-1,1)$$

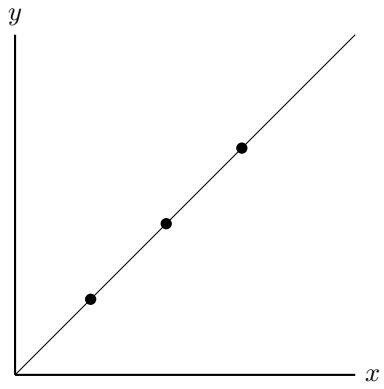
   If our goal is to minimize reconstruction error, which eigenvalue/eigenvector pair would we choose?
   Solution.

   $$\lambda_1, v_1$$

   (c) What is the reconstruction error for the first principal component?
   Solution.

   $$0$$

   (d) Create an x/y axis and draw the data points and the principal component from part (c).
   Solution.



Recommender Systems

1. (a) When doing matrix factorization, how do we find the best k latent factors?

   Solution. Hold out some of the input ratings as a validation set and try multiple values of k, choosing the one that minimizes validation error.

   (b) Consider the case of learning a featurized matrix factorization model. Suppose instead of using a linear model we use a neural network. Is it possible to use a neural network like this? If not, explain why. If it is possible, give one benefit and drawback of using a neural network here.

   Solution. Yes it is possible. Neural networks have the advantage of being able to learn more complex relationships than a linear model. However, they require more training data since otherwise they are likely to overfit.

2. (a) Do recommendation systems fit into supervised or unsupervised learning?

   Solution. There are aspects of both supervised and unsupervised learning. Like supervised learning, there are known outputs (i.e. items that a user purchases). Yet they are similar to unsupervised learning in that we want to find structure and similarity between users and items.

   (b) Recall the matrix of users and ratings with missing data from class. Denote the (u, v)-th entry of this matrix as $r_{u,v}$, the user specific weights as L_u and rating specific weights as R_v.

   i. In matrix factorization, how is $r_{u,v}$ modeled in terms of L_u, R_v?

   Solution.

   $$r_{u,v} = L_u \cdot R_v$$

   ii. What is the quality metric that is used to estimate L_u, R_v?

   Solution.

   $$\sum_{u,v:r_{u,v}\neq ?} (L_u \cdot R_v - r_{uv})^2$$

   iii. If your movie ratings were between 1 and 5, is it possible for the approximate ratings found by minimizing the above quality metric to be above 5?

   Solution. Yes.