Rank of a Matrix

- Dimension of vector space spanned by columns of A
- Simple algorithm to determine rank of a matrix: compute the row reduction and count the number of pivots
- DM or ML our data is ultimately represented by a matrix so we can compute its rank
- An underlying *structure* might be the fact that the data matrix has a smaller rank than the dimensionality of the feature space

Projection Operator onto a Subspace

- Dimension of vector space spanned by columns of A
- The columns of A are an orthonormal basis for the subspace that we want to project onto

$$P = AA^T \tag{1}$$

- Projection operators are *idempotent*

$$P^2 = P \tag{2}$$

Recap

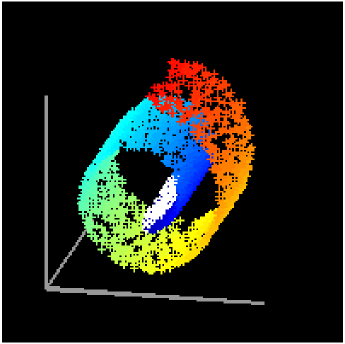
- Feature selection and data preprocessing is an essential consideration in a learning context
- Transform the data (vectors in the feature space) such that the values are meaningful for our problem and ideally only have a few independent values
- Methods for evaluating features
  - Dimensionality reduction techniques such as PCA
  - Use standard statistical hypothesis tests to find features that significantly vary over the classes
  - Fix an algorithm and quantify the success of predictions using different feature sets
  - Exploit domain knowledge

Other Feature Selection Approaches

- Generally dealing with high-dimensional vector spaces
- However, we would like to carry over some of our low dimensional intuitions
- One notion of spaces that seems to be well understood is of one space being embedded in another
- Example: Imagine the surface of a ball, our ideas about the *shape* of the ball somehow rely on our understanding that it is a two-dimensional space embedded in the three dimensional world (this is what we call a *surface*)

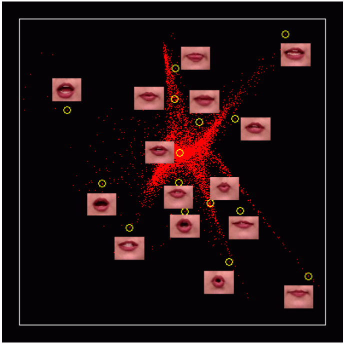
Manifold Learning

- In the same way, when we perform dimensionality reduction or try to find an optimal subset features we assume that our data is somehow embedded in a lower dimensional subspace
- One approach to finding a non-linear embedding is called *locally linear embedding* (Roweis *et al*)



Locally Linear Embedding (example)

- The LLE approach can be used to find the underlying structure in a set of natural images [https://www.cs.nyu.edu/ roweis/lle/]



Readings

- Read these papers over the next week
- Guyon, Isabelle, and Andr Elisseeff. "An introduction to variable and feature selection." The Journal of Machine Learning Research 3 (2003): 1157-1182.
- Saul, Lawrence K., and Sam T. Roweis. "An introduction to locally linear embedding." unpublished. Available at: http://www. cs. toronto. edu/ roweis/lle/publications. html (2000).

Ridge Regression and LASSO

- Two common means of regularizing least squares are called ridge regression and lasso
- Ridge regression: $J(h) = ||y - \mathbf{X}h||_2 + \lambda \sum_k h_k^2$
- LASSO: $J(h) = ||y - \mathbf{X}h||_2 + \lambda \sum_k |h_k|$
- Consider what happens as the parameter λ is varied from 0 to infinity

Lasso Regression in Python

```
lasmodel = linear_model.Lasso(alpha=0.01,fit_intercept=False)
lasmodel.fit(tt.values[:,0:4],tt.values[:,4])

print(lasmodel.coef_)
```

- Import `linear_model` from scikit learn
- The alpha argument determines the strength of the regularization term

Task

- Download the lasso_example dataset from the course moodle
- Apply regularized linear regression as shown in the slides to the dataset while varying the value of the regularization coefficient
- Create a plot that shows how the model fits the data
- Also plot the results as a function of the alpha parameter