HW 3: Unsupervised Learning

16:198:461

In this assignment, we want to take a data set and extract some measure of structure from it.

Let $X_1 \sim \mathcal{N}(0, 1)$. Define

$$\begin{aligned} X_4 &= X_1 + \mathcal{N}(0, \sigma^2), \\ X_7 &= X_4 + \mathcal{N}(0, \sigma^2), \\ X_{10} &= X_7 + \mathcal{N}(0, \sigma^2), \\ X_{13} &= X_{10} + \mathcal{N}(0, \sigma^2), \\ &\ldots \\ X_{28} &= X_{25} + \mathcal{N}(0, \sigma^2), \end{aligned} \tag{1}$$

and

$$\begin{aligned} X_2 &= X_1 + \mathcal{N}(0, \sigma^2), \\ X_5 &= X_2 + \mathcal{N}(0, \sigma^2), \\ X_8 &= X_5 + \mathcal{N}(0, \sigma^2), \\ X_{11} &= X_8 + \mathcal{N}(0, \sigma^2), \\ &\ldots \\ X_{29} &= X_{26} + \mathcal{N}(0, \sigma^2), \end{aligned} \tag{2}$$

and

$$\begin{aligned} X_3 &= X_1 + \mathcal{N}(0, \sigma^2), \\ X_6 &= X_3 + \mathcal{N}(0, \sigma^2), \\ X_9 &= X_6 + \mathcal{N}(0, \sigma^2), \\ X_{12} &= X_9 + \mathcal{N}(0, \sigma^2), \\ &\ldots \\ X_{30} &= X_{27} + \mathcal{N}(0, \sigma^2). \end{aligned} \tag{3}$$

Generate a data set based on the above, with 5000 data points, using $\sigma^2 = 0.1$ (at first).

Problem 1: Auto-Encoders

We can define an auto-encoder, a neural network mapping from the input space to the input space. A simple model maps the input vector $\underline{x}$ to a hidden layer of k nodes, to an output layer that matches the input layer. We can take the hidden layer as having an activation function of $\sigma(z) = (e^z - e^{-z})/(e^z + e^{-z})$, aka $\tanh(z)$, and the output layer as having activations of $\sigma(z) = z$, so as to produce real valued outputs with no bounds. Building a network $F(\underline{x})$ in this way, we want to minimize the loss function

$$\frac{1}{N} \sum_{i=1}^{N} ||\underline{x}^i - F(\underline{x}^i)||^2 \tag{4}$$

For values of k between 1 and 30, train an auto encoder on the data, and plot the final loss as a function of k. What does this suggest about the dimensionality of the data set?

How does this change as σ^2 changes, between 0 and 2? Why?

Problem 2: PCA

For the data matrix X, compute the the principal components of the data by computing the eigenvalues and vectors of $X^T X$. What does the results suggest about the dimensionality of the data set? Is this result robust (do you get the same result doing it again on a different data set)?

How does the result change as σ^2 changes, between 0 and 2?

Problem 3: Correlation Graphs

For each $i = 1, 2, \ldots, 30$, consider the problem of predicting the value of X_i from the other features. By fitting a model to predict X_i from the other features, we can identify what other features are most predictive of X_i, i.e., what other features X_i most depends on. Consider doing so with this data set and linear models, and building a dependency graph between the features by connecting a feature X_i to the features it most depends on.

Consider building this graph by connecting each feature to the two other features it most depends on in each model. Does this graph reproduce the 'true' dependency structure of this data? Is this robust (do you get the same result doing it again on a different data set)? What if you consider connecting each feature to the three most predictive features, or four most?

Consider building this graph by connecting each feature to the other features that have a weight in the model larger than some threshold. How does the resulting graph depend on the threshold taken? Are you able to reconstruct the true dependency graph?

How does the result change as σ^2 changes, between 0 and 2?