Metric Learning

Goal: learn how A and B sections of a particular tune are related.

- Learn a metric $d(x, y)$ so that if $x, y \in \mathbb{R}^n$ correspond to the A and B sections of a tune, $d(x, y)$ is small.
- Commonly used: the Mahalanobis metric, given by $d(x, y) = \|x - y\|_M = \sqrt{(x - y)^T M (x - y)}$ for M positive semidefinite.

Such so-called "metric learning" problems usually also have pairs of points intended to be dissimilar, e.g. given pairs of similar points $(x^{(i)}, x^{(j)}) \in \mathcal{S}$ and pairs of dissimilar points $(y^{(i)}, y^{(j)}) \in \mathcal{D}$, solve the following problem for M:

$$
\begin{array}{ll}
\text{minimize} & \sum_{(x^{(i)}, x^{(j)}) \in \mathcal{S}} \left\| x^{(i)} - x^{(j)} \right\|_M^2 \\
\text{subject to} & \sum_{(y^{(i)}, y^{(j)}) \in \mathcal{D}} \left\| y^{(i)} - y^{(j)} \right\|_M^2 \geq 1 \qquad [1]. \\
& M \succeq 0
\end{array}
$$

However, it is unclear how to specify dissimilar pairs of A and B sections.

- Without the dissimilarity constraint, the trivial solution is $M = 0$.
- We add the convex constraint $(\det M)^{1/n} \geq 1$ which implies $\det M = 1$:

$$
\begin{array}{ll}
\text{minimize} & \displaystyle\sum_{k=1}^{m} \left\| x^{(k)} - y^{(k)} \right\|_M^2 \\
\text{subject to} & M \succeq 0 \\
& (\det M)^{1/n} \geq 1.
\end{array}
$$

Measuring success of our learned metric:

(1) We use 8-fold cross-validation.

(2) Fit a 35-component PCA model to our training set and, using this model, solve for a 35×35 M (and therefore the metric $d(x, y)$).

(3) Match B sections to given A sections (or vice-versa): for each A section in the training set, order all B sections in the training set by distance from the A section. Record the rank of the correct B section (first is optimal).