Maximum Variance Unfolding

Similarly to other manifold learning techniques like Isomap and Locally Linear Embedding, Maximum Variance Unfolding (MVU) constructs a graph representation of the data points. Its goal is to unfold the manifold in a lower-dimensional subspace by maximizing the variance of the projected data while maintaining local isometry for each point. Given the original data $\{\boldsymbol{x}_i\}_{i=1}^n \in \mathbb{R}^D$, we write the MVU problem:

$$\begin{array}{ll} \max\limits_{\boldsymbol{y}_1,...,\boldsymbol{y}_n} & \sum\limits_{i=1}^n \|\boldsymbol{y}_i\|_2^2 \\ \text{s.t.} & \|\boldsymbol{y}_k - \boldsymbol{y}_l\|_2^2 = \|\boldsymbol{x}_k - \boldsymbol{x}_l\|_2^2, \quad k \sim l \\ & \sum\limits_{i=1}^n \boldsymbol{y}_i = \boldsymbol{0} \end{array}$$

We have the following:

- The objective function maximizes the variance of the embedded data $\boldsymbol{y}$.
- The first constraint represents our wish to preserve local isometry, i.e., distances of each point to its k-nearest neighbors should be the same in the embedding space. ($k \sim l$ represents a nearest-neighbor relationship.)
- In order to avoid degenerate solutions, we zero-center the data in the last constraint.

Rewriting the problem:

$$\begin{array}{ll} \max\limits_{\boldsymbol{y}_1,...,\boldsymbol{y}_n} & \sum\limits_{i=1}^n \boldsymbol{y}_i^\top \boldsymbol{y}_i \\ \text{s.t.} & \boldsymbol{y}_k^\top \boldsymbol{y}_k - 2\boldsymbol{y}_k^\top \boldsymbol{y}_l + \boldsymbol{y}_l^\top \boldsymbol{y}_l = \|\boldsymbol{x}_k - \boldsymbol{x}_l\|_2^2, \quad k \sim l \\ & \sum\limits_{i=1}^n \boldsymbol{y}_i = \boldsymbol{0} \end{array}$$

It is clear that this problem is not convex:

- The objective function is quadratic in the variable of interest, so maximizing (as opposed to minimizing it) it is not a convex problem.
- The first constraint is defined in terms of bilinear terms $\boldsymbol{y}_k^\top \boldsymbol{y}_l$.

If we collect the target points into the columns of a matrix $\boldsymbol{Y} = [\boldsymbol{y}_1 \quad \cdots \quad \boldsymbol{y}_n]$, we can rewrite the problem:

$$\begin{array}{ll} \max\limits_{\boldsymbol{Y}} & \operatorname{tr}(\boldsymbol{Y}^\top \boldsymbol{Y}) \\ \text{s.t.} & \boldsymbol{e}_k^\top \boldsymbol{Y}^\top \boldsymbol{Y} \boldsymbol{e}_k - 2\boldsymbol{e}_k^\top \boldsymbol{Y}^\top \boldsymbol{Y} \boldsymbol{e}_l + \boldsymbol{e}_l^\top \boldsymbol{Y}^\top \boldsymbol{Y} \boldsymbol{e}_l = \|\boldsymbol{x}_k - \boldsymbol{x}_l\|_2^2, \quad k \sim l \\ & \boldsymbol{1}^\top \boldsymbol{Y}^\top \boldsymbol{Y} \boldsymbol{1} = 0, \end{array}$$

where $\boldsymbol{e}_k$ is a one-hot vector with 1 at index k, i.e., can be thought of as a "selection vector". Note that now, the whole problem depends on $\boldsymbol{Y}^\top \boldsymbol{Y}$, so we can introduce a new variable $\boldsymbol{G} = \boldsymbol{Y}^\top \boldsymbol{Y}$ to "linearize" the terms that depend on $\boldsymbol{Y}^\top \boldsymbol{Y}$:

$$\begin{array}{ll} \max\limits_{\boldsymbol{G},\boldsymbol{Y}} & \operatorname{tr}(\boldsymbol{G}) \\ \text{s.t.} & \boldsymbol{e}_k^\top \boldsymbol{G} \boldsymbol{e}_k - 2\boldsymbol{e}_k^\top \boldsymbol{G} \boldsymbol{e}_l + \boldsymbol{e}_l^\top \boldsymbol{G} \boldsymbol{e}_l = \|\boldsymbol{x}_k - \boldsymbol{x}_l\|_2^2, \quad k \sim l \\ & \boldsymbol{1}^\top \boldsymbol{G} \boldsymbol{1} = 0 \\ & \boldsymbol{G} = \boldsymbol{Y}^\top \boldsymbol{Y} \end{array}$$

Because $\boldsymbol{G} = \boldsymbol{Y}^\top \boldsymbol{Y}$ defines an inner product matrix, we can replace that constraint if we make sure that $\boldsymbol{G}$ is both symmetric, positive semidefinite and that the rank of $\boldsymbol{G}$ is greater than the dimension of the $\boldsymbol{y}$s:

$$
\begin{aligned}
\max_{\boldsymbol{G}} \quad & \operatorname{tr}(\boldsymbol{G}) \\
\text{s.t.} \quad & \boldsymbol{e}_k^\top \boldsymbol{G} \boldsymbol{e}_k - 2\boldsymbol{e}_k^\top \boldsymbol{G} \boldsymbol{e}_l + \boldsymbol{e}_l^\top \boldsymbol{G} \boldsymbol{e}_l = \|\boldsymbol{x}_k - \boldsymbol{x}_l\|_2^2, \quad k \sim l \\
& \boldsymbol{1}^\top \boldsymbol{G} \boldsymbol{1} = 0 \\
& \boldsymbol{G} \succeq 0 \\
& \operatorname{rk}(\boldsymbol{G}) \leq n
\end{aligned}
$$

Now, the only nonconvexity in the problem is given by the rank constraint. If we remove it, we arrive at a convex relaxation of the original problem, which is the final formulation of MVU:

$$
\begin{aligned}
\max_{\boldsymbol{G}} \quad & \operatorname{tr}(\boldsymbol{G}) \\
\text{s.t.} \quad & \boldsymbol{e}_k^\top \boldsymbol{G} \boldsymbol{e}_k - 2\boldsymbol{e}_k^\top \boldsymbol{G} \boldsymbol{e}_l + \boldsymbol{e}_l^\top \boldsymbol{G} \boldsymbol{e}_l = \|\boldsymbol{x}_k - \boldsymbol{x}_l\|_2^2, \quad k \sim l \\
& \boldsymbol{1}^\top \boldsymbol{G} \boldsymbol{1} = 0 \\
& \boldsymbol{G} \succeq 0
\end{aligned}
$$

Here, $\boldsymbol{G} \in \mathbb{R}^{(n \times n)}$ is a Gramian (or inner product) matrix, so that maximizing its trace corresponds to maximizing the variance of the data in the target space.