Randomization, Quantification and Visualization

Myung-Hoe Huh

Abstract — I will present three topics of my research interests: Randomization, Quantification and Visualization. First, I report the lack-of-randomness in shuffling of HwaTu or Hanafuda cards (Huh and Lee, 2010). Second, I write a multidimensional scaling procedure for asymmetric distance matrices (Huh and Lee, 2011). Lastly, nonparametric classifiers produced by support vector machine are visualized in reduced dimensions (Huh and Park, 2010).

Keyword: Card shuffling, Asymmetric MDS, SVM visualization.

1 Hanafuda Card Shuffling

Physically randomizing the sequence of cards is never an easy task. For instance, US 1970 draft lottery resulted in a social turmoil since the outcome sequence of birthday numbers was similar to the input order.

For rifle shuffling of 52 game cards, Bayer and Diaconis (1992) showed that common practices produce poor randomness and argued that at least seven replications are necessary for fairly good randomness. In this section, I propose a lottery box model for random ordering of $n$ cards and identify basic properties.

For HwaTu or Hanafuda card plays which are very popular in Korea, the shuffling procedure can be described as follows.

1) $48 (= n)$ cards are split into $s$ blocks of random sizes. Here $s$ can be varied from person to person, ranging from 3 to 9. We assume that block sizes are realized from Multinomial $(n, \mu)$, where $\mu = \frac{1}{s}(1, \ldots, 1)$.

2) The orders of blocks are reversed: from 1, $\cdots$, $s$ to $s$, $\cdots$, 1.

3) Repeat Step 1 ad Step 2 $m$ times.

Since transitions are Markovian and doubly stochastic, the limit of repetitions must be the steady state of complete randomness. Remaining question is how large $m$ should be?

For perfectly random permutations, we may expect that the correlation with the starting permutation $(1, \cdots, 48)$ is distributed to

\[
\begin{array}{cccccccc}
1\% & 5\% & 10\% & 25\% & 50\% & 75\% & 90\% & 95\% & 99\% \\
-0.337 & -0.241 & -0.188 & -0.101 & -0.001 & 0.099 & 0.187 & 0.240 & 0.336.
\end{array}
\]

Also, the occurrence of consecutive cards in perfectly random permutations

\[
Z = \sum_{j=1}^{n-1} I(X_{j+1} = (X_j \mod n) + 1)
\]

should show two characteristics: 1) $E(Z) = 1$, and 2) $Z$ follows Poisson(1) approximately. However, we found the following facts through Monte-Carlo experiments.

*Professor, Dept. of Statistics, Korea University. Seoul 136-701, Korea. E-mail: stat420@korea.ac.kr
Study 1: \( s = 4, m = 6, 12, 24, 48, 96, 144, 192 \).

<table>
<thead>
<tr>
<th>Median Corr.</th>
<th>Mean of Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s = 4, m = 6 ):</td>
<td>0.509, 32.7</td>
</tr>
<tr>
<td>( s = 4, m = 12 ):</td>
<td>0.259, 23.2</td>
</tr>
<tr>
<td>( s = 4, m = 24 ):</td>
<td>0.069, 11.7</td>
</tr>
<tr>
<td>( s = 4, m = 48 ):</td>
<td>-0.006, 3.6</td>
</tr>
<tr>
<td>( s = 4, m = 96 ):</td>
<td>-0.002, 1.2</td>
</tr>
<tr>
<td>( s = 4, m = 144 ):</td>
<td>0.007, 1.0</td>
</tr>
<tr>
<td>( s = 4, m = 192 ):</td>
<td>-0.007, 1.0</td>
</tr>
</tbody>
</table>

Study 2: \( s = 8, m = 6, 12, 24, 48, 96, 144, 192 \).

<table>
<thead>
<tr>
<th>Median Corr.</th>
<th>Mean of Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s = 8, m = 6 ):</td>
<td>0.788, 19.0</td>
</tr>
<tr>
<td>( s = 8, m = 12 ):</td>
<td>0.686, 8.5</td>
</tr>
<tr>
<td>( s = 8, m = 24 ):</td>
<td>0.404, 2.4</td>
</tr>
<tr>
<td>( s = 8, m = 48 ):</td>
<td>0.158, 1.0</td>
</tr>
<tr>
<td>( s = 8, m = 96 ):</td>
<td>0.027, 1.0</td>
</tr>
<tr>
<td>( s = 8, m = 144 ):</td>
<td>0.005, 1.0</td>
</tr>
<tr>
<td>( s = 8, m = 192 ):</td>
<td>-0.008, 1.0</td>
</tr>
</tbody>
</table>

Hence, in order to obtain "random" permutations, we should repeat the shuffling at least 96 times, for \( s \) between 4 and 8. But, in reality, typical Hanafuda players shuffle around 6 (\( = m \)) times only. The consequence is the increased opportunities or risks, incurred due to poor randomness. We should guard ourselves against the false assumption of randomness in playing Hanafuda games.

2 Asymmetric MDS

In most cases of multidimensional scaling (MDS), the distances or dissimilarities among units are assumed to be symmetric. Asymmetric MDS developed so far are very complex (Chino, 2011), so that nonexpert users often feel helpless. I propose the following algorithm for asymmetric MDS, which is simple in concept and in computation.

Denote \( n \times n \) asymmetric distance matrix by \( D (= (d_{ij})) \). Define

\[
e_{ij} = d_{ij} - d_{ji}, \quad i, j = 1, \ldots, n.
\]

Thus \( E (= (e_{ij})) = D^T - D \). \( e_{ij} > 0 \) means that \( d_{ij} < d_{ji} \), so that in mountain hiking the point \( i \) is located higher than the point \( j \) in terms of difficulty or time.

Suppose that the point \( i \)'s quantified "altitude" is \( z_i \). Is it possible to obtain quantified altitudes \( z_1, \ldots, z_n \) such that \( z_i > z_j \) if \( e_{ij} > 0 \) and that \( z_i < z_j \) otherwise? The answer is "no", but we may formulate the task as follows.

\[
\max_{z_1, \ldots, z_n} \sum_{i=1}^{n} \sum_{j=1}^{n} e_{ij}(z_i - z_j)
\]

subject to \( \sum_{i=1}^{n} z_i^2 = 1 \) and \( \sum_{i=1}^{n} z_i = 0 \).

The solution is

\[
z = \frac{e^{[1]}}{||e^{[1]}||}
\]
where \( e^{ii} = (e_{i1}, \ldots, e_{in}) \), \( e_{i} = \sum_{j=1}^{n} e_{ij} \).

**Example: Cross-referencing among Psychology Journals**

Table 1 lists the cross-referencing frequencies among eight psychology journals. Element \( a_{ij} \) is the number of the journal \( j \) citations in the journal \( i \). Thus, \( a_{ij} \)'s are transformed to distances \( d_{ij} \)'s by

\[
\frac{1}{d_{ij}} = 1 + a_{ij}.
\]

Following numbers representing the "altitudes" are obtained by the quantification procedure proposed above.

<table>
<thead>
<tr>
<th>Journals</th>
<th>AJP</th>
<th>JAS</th>
<th>JAP</th>
<th>CPP</th>
<th>JCP</th>
<th>JED</th>
<th>JEX</th>
<th>PKA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altitudes</td>
<td>-0.32</td>
<td>0.05</td>
<td>0.13</td>
<td>0.00</td>
<td>0.68</td>
<td>0.23</td>
<td>-0.57</td>
<td>-0.19</td>
</tr>
</tbody>
</table>

JEX is the lowest, while JCP is the highest. It means that JEX quotes other journals heavily and that JCP is being quoted intensively by other journals. Figure 1 is a metric MDS of the harmonic average of two distance matrices, \( D \) and \( D' \), being colored from red, orange, yellow, green to blue by the altitude, from the highest to the lowest.

**Example: Knoke's Information Network**


Table 2 lists the shortest distances between institutions, which are asymmetric. For instance, the distance from INDU to COUN [1] is 1, while the distance form COUN to INDU [4] is 2.

The following numbers are quantification of the altitudes for ten institutions. See Figure 2 for a nonmetric MDS colored by the altitude. NEWS(Newspaper) is the lowest (-0.64), meaning that NEWS absorbs the information to the maximum degree.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Altitudes</td>
<td>-0.09</td>
<td>-0.09</td>
<td>0.18</td>
<td>-0.18</td>
<td>0.00</td>
<td>0.46</td>
<td>-0.64</td>
<td>0.37</td>
<td>-0.28</td>
<td>0.28</td>
</tr>
</tbody>
</table>
3 Visualizing SVM Classifiers

Support vector machines (SVMs) are known as flexible and efficient classifier of multivariate observations, producing a hyperplane or hyperdimensional curved surface in multidimensional feature space that best separates training samples by known groups. As various methodological extensions are made for SVM classifiers in recent years, it becomes more difficult to understand the constructed model intuitively. The aim of this section is to visualize various SVM classifications tuned by several parameters in reduced dimensions, so that data analysts secure the tangible image of the products that the machine made.

Suppose that we have $n$ observations $(x_1, y_1), \ldots, (x_n, y_n)$ each of which belongs to one of two classes, Class 1 or Class 0, where $x_i$'s are $p \times 1$ vectors of feature variables on the continuous scale and $y_i$'s are either 1 (for Class 1) or -1 (for Class 0) according to the associated group.

Table 2: Knoke's Information Network: Shortest Distances

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>COUN</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>COMM</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>EDUC</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>INDU</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>MAYR</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>WRO</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>NEWS</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>UWAY</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>WELF</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>WEST</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 1: MDS of Psychology Journals colored by the altitude
Visualizing Linear SVM classifications

In linear SVM classifications, the classification boundary is specified as \( f(x) = 0 \), or
\[
wx + b = 0,
\]
which is perpendicular to the unit vector \( w/\|w\| = v[1] \) in \( p \)-dimensional feature space. Hence, to visualize \( n \) observed units with predicted class labels effectively, I propose to plot dots at \( v[1]x_1, \ldots, v[1]x_n \) in different colors depending on the predicted class ("blue" for Class 1 and "red" for Class 0). Since two-dimensional graph can be drawn without much additional effort, we add the second axis, which is orthogonal to the first axis determined by \( w \) or \( v[1] \). Therefore we proceed as follows.

First, project \( p \times 1 \) standardized vectors \( x_1, \ldots, x_n \) on the directional vector \( v[1] \) and compute residual vectors \( x^{[1]}_1, \ldots, x^{[1]}_n \). That is,
\[
x^{[1]}_i = x_i - v[1]v[1]x_i, \quad i = 1, \ldots, n.
\]
Second, compute the principal component direction vector \( v[2] \) that carries maximally dispersed projections of \( x^{[1]}_1, \ldots, x^{[1]}_n \). That is, \( v[2] \) is the eigenvector corresponding to the largest eigenvalue of \( X^{[1]} X^{[1]} \), where \( X^{[1]} = (x^{[1]}_1, \ldots, x^{[1]}_n) \).

Third, \( n \) observations are dotted at
\[
\]
on the first and second axis, respectively, in different colors depending on their membership class ("blue" for Class 1 or "red" for Class 0). Accordingly, on the first and second axis, the feature variables are plotted with arrow heads at \( p \) components of
\[
\]
On the 2D projection plane, we may superimpose probability contours as follows. For \( u' = (u_1, u_2) \) of the 2D projection plane, attach to \( x = u_1v[1] + u_2v[2] \) the probability \( p_y(1) \) being classified to Class 1. Then, we add the probability contours to the 2D display of SVM classification.
Visualizing Nonlinear SVM classifications

In nonlinear SVM classifications, the classification function is given as

$$ f_N(x) = \sum_{i=1}^{n} \lambda_i y_i K(x_i, x) + b. $$

Hence the gradient of $f_N(x)$ can be written as

$$ \nabla f_N(x) = \sum_{i=1}^{n} \lambda_i y_i \nabla K(x_i, x). $$

For the Gaussian RBF $K(x, x') = \exp(-\gamma \|x - x'|^2)$,

$$ \nabla f_N(x) = -2 \gamma \sum_{i=1}^{n} \lambda_i y_i \exp(-\gamma \|x_i - x\|^2)(x_i - x). $$

Since constructed SVM classification depends only on $L \leq n$ support vectors $X_s = \{x_{s_1}, x_{s_2}, \ldots, x_{s_L}\}$, we restrict our attention to the gradient vectors at $X_s$: For Gaussian RBF,

$$ \nabla f_N(x_s) = -2 \gamma \sum_{i=1}^{n} \lambda_i y_i \exp(-\gamma \|x_i - x_s\|^2)(x_i - x_s), \quad l = 1, \ldots, L. $$

We propose the 2D display for nonlinear SVM classifications as follows.

First, obtain $p \times L$ matrix $G'$ consisting of gradient vectors as columns:

$$ G' = (\nabla f_N(x_{s_1}), \ldots, \nabla f_N(x_{s_L})). $$

Second, compute principal component direction vectors $v^{[1]}$ and $v^{[2]}$ that carries maximally dispersed projections of $\nabla f_N(x_{s_1}), \ldots, \nabla f_N(x_{s_L})$. That is, $v^{[1]}$ and $v^{[2]}$ are the eigenvectors corresponding to the largest and second largest eigenvalues of $G'G$. Thus $\|v^{[1]}\| = 1, \|v^{[2]}\| = 1, v^{[1]}v^{[2]} = 0$.

Third, $n$ observations are dotted at $z_i^{[1]} = v^{[1]}x_i, z_i^{[2]} = v^{[2]}x_i$ on the first and the second axis, respectively, colored differently depending on their membership class ("red" for Class 0 or "blue" for Class 1). Accordingly, on the first and second axis, the feature variables are plotted with arrow heads at $p$ components of $y^{[1]} = v^{[1]}, y^{[2]} = v^{[2]}$.

Fourth, we superimpose probability contours as in the case of linear SVM classifications.

$m$-class Problem

For classification problem for more than $m \geq 2$ classes, SVM classification consists of all pairwise classifications: Inside the $m$-class SVM classification, there exist $mC_2$ classification functions of the form $f^{[k_1, k_2]}(x)$, which is expressed with $L_{[k_1, k_2]}(\leq n)$ support vectors, that distinguishes Class $k_1$ against Class $k_2$. Thus we need to gather the gradient vectors into the matrix $G_{k_1, k_2}$ of $f^{[k_1, k_2]}(x)$ for $1 \leq k_1 < k_2 \leq m$ at the support vectors for respective classification, where

$$ G_{k_1, k_2} = (\nabla f^{[k_1, k_2]}(x_{s_1}), \ldots, \nabla f^{[k_1, k_2]}(x_{s_{L(k_1, k_2)}})). $$
Figure 3: Visualization of the linear SVM classification for Italian Olive Oils data from Four Regions. Dark solid curve is the classification boundary. Dotted curves denote probability contours at the 0.9 level.

and join all $G^l_{k_1,k_2}$'s side by side:

$$G^l_{k_1,k_2} = (G^l_{1,2},\ldots,G^l_{1,m},\ldots,G^l_{m-1,m}).$$

Then, compute principal component direction vectors $v^{[1]}$ and $v^{[2]}$ that carry maximally dispersed projections of all column vectors of $G^l$. That is, $v^{[1]}$ and $v^{[2]}$ are the eigenvectors corresponding to the largest and second largest eigenvalues of $G^l G$. Thus $\|v^{[1]}\| = 1, \|v^{[2]}\| = 1, v^{[1]} v^{[2]} = 0$.

Hence, $n$ observations are dotted at

$$z^{[1]}_i = v^{[1]} x_i, z^{[2]}_i = v^{[2]} x_i$$

on the first and second axis, respectively, colored differently depending on their membership class ("blue" for Class 1, "green" for Class 2,\ldots, "red" for Class $m$). Accordingly, on the first and second axis, the feature variables are plotted with arrow heads at $p$ components of

$$y^{[1]} = v^{[1]}, y^{[2]} = v^{[2]}.$$ 

Example: Olive Oils data from four South areas of Italy

The Olive Oils data consists of eight fatty acid composition measurements (palmitic, palmitoleic, stearic, oleic, linoleic, linolenic, arachidic, and eicosenoic) in three regions (North, South, Sardina), which can be divided into nine areas. In this example, we restrict our attention to the data from South region which is divided into four areas (Calabria, North-Apulia, Sicily, South-Apulia). Thus the number of classes is four. Figure 3 shows the classification map of a linear SVM classification.
References


