Research notes
Information Theory
Gowtham Kumar

Document #2
Sunday, July 25, 2010

On Sequences of Pairs of Dependent Random Variables
A simpler proof of the main result using SVD

1 A simple optimization problem

Consider the following optimization:

$$\begin{aligned}&\text{maximize } x^T My\\&\text{subject to}\\&||x|| = 1, \quad ||y|| = 1\end{aligned}$$

where $M \in R^{m\times n}$, $x \in R^m$ and $y \in R^n$. For the case when M is a diagonal matrix,

$$\begin{aligned}x^T My &= \sum M_{ii} x_i y_i\\&\leq \left(\max_i |M_{ii}|\right) x^T y\\&\leq \left(\max_i |M_{ii}|\right) ||x||.||y||\\&= \left(\max_i |M_{ii}|\right)\end{aligned}$$

Equality can be achieved by setting $x_i = y_i = 1$ and $x_j = y_j = 0$ for $j \neq i$ For any general matrix $M \in R^{m\times n}$ of rank r, singular value decomposition yields

$$M = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

where $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_r$ are the singular values, $u_1, u_2, \ldots$ are the left singular vectors and $v_1, v_2, \ldots$ are the right singular vectors. In that case, using u_i's as a basis for x and v_i's as a basis for y, any general matrix M is replaced by a diagonal matrix Σ of singular values. Thus the solution to the optimization problem is the largest singular value σ_1 with $x^* = u_1$ and $y^* = v_1$.

2 Tensor product and singular values

Let $A \in R^{m\times n}$ and $B \in R^{m_1\times n_1}$ be any 2 matrices. The tensor product of A and B is defined as:

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & . & . & a_{1n}B \\ a_{21}B & a_{22}B & . & . & a_{2n}B \\ . & . & . & . & . \\ . & . & . & . & . \\ a_{m1}B & a_{m2}B & . & . & a_{mn}B \end{pmatrix}$$

where a_{ij} is the (i,j)th element of the matrix A. Thus $A \otimes B \in R^{mm_1\times nn_1}$. We state a lemma without proof.

Lemma: If

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

$$B = \sum_{j=1}^{s} \lambda_j w_j z_j^T$$

represent the SVD of A and B respectively, then

$$A \otimes B = \sum_{i=1}^{r}\sum_{j=1}^{s} \sigma_i\lambda_j (u_i \otimes w_j)(v_i \otimes z_j)^T$$

Refer to wikipedia for more details.

3 Renyi Correlation

Let $(X,Y) \sim p(x,y)$ be a pair of dependent RV's. Then the Renyi Correlation $\rho(X,Y)$ is defined as

$$\rho(X,Y) = \max_{f(x),g(y)} Ef(X)g(Y)$$
$$\text{subject to}$$
$$Ef(X) = Eg(Y) = 0$$
$$Ef(X)^2 = Eg(Y)^2 = 1$$

where $f(x)$ and $g(y)$ are functions of x and y respectively.

Computation of Renyi correlation can be thought of as an optimization problem. Let $p(x,y)$ be the joint distribution of (X,Y) and let $p_X(x), p_Y(y)$ be it's marginals. Define $u(x) = f(x)\sqrt{p_X(x)}$ and $v(y) = \sqrt{p_Y(y)}$. Then

*g(y) (typo)

$$\rho(X,Y) = \sum_{x,y} u(x)v(y)\frac{p(x,y)}{\sqrt{p_X(x)}\sqrt{p_Y(y)}} = u^t Qv$$

not RhoMax

where $Q(x,y) = \frac{p(x,y)}{\sqrt{p_X(x)}\sqrt{p_Y(y)}}$ are the elements of the matrix Q

Using this change of variables, the problem of finding Renyi correlation becomes

$$\rho(X,Y) = \max_{u,v} u^T Qv$$

Best u,v yields Rhomax

also note, Quantizing u,v doesn't quantize f,g

subject to

$$\sum_x u(x)\sqrt{p_X(x)} = \sum_y v(y)\sqrt{p_Y(y)} = 0$$

=E(f(x))=E(g(y))=0

$$||u|| = ||v|| = 1$$

If the zero mean constraint on f, g were absent, the solution to the optimization would be to choose $f(x) = g(y) = 1$ to obtain $Ef(X)g(Y) = 1$, which is the largest correlation coefficient permissible and the largest singular value of the matrices P, Q. Due to the zero mean constraints, we should pick $u(x)$ orthogonal to the subspace $\sqrt{p_X(x)}$ and similarly for $v(y)$. However, the singular vectors are orthogonal. Hence Renyi Correlation is the second largest singular value of Q and u, v are the corresponding left, right singular vectors respectively. Thus $\rho(X,Y) = \sigma_2(Q)$.

4 Main Theorem

Theorem: Let $(X_1, Y_1) \sim p_1(x_1, y_1)$ and $(X_2, Y_2) \sim p_2(x_2, y_2)$ be independent pairs of RVs. Then, $\rho(X_1X_2, Y_1Y_2) = \max(\rho(X_1, Y_1), \rho(X_2, Y_2))$.

Proof: The proof can be obtained by combining the results of sections 2 and 3. From the probability matrix P, we obtained the probability matrix Q. Note that, using the independence of the pairs,

$$\begin{aligned}
Q(x_1, x_2, y_1, y_2) &= \frac{P(x_1, x_2, y_1, y_2)}{\sqrt{P_{X_1,X_2}(x_1, x_2)}\sqrt{P_{Y_1,Y_2}(y_1, y_2)}} \\
&= \frac{p_1(x_1, y_1)}{\sqrt{p_{X_1}(x_1)}\sqrt{p_{Y_1}(y_1)}} \frac{p_2(x_2, y_2)}{\sqrt{p_{X_2}(x_2)}\sqrt{p_{Y_2}(y_2)}} \\
&= Q_1(x_1, y_1)Q_2(x_2, y_2)
\end{aligned}$$

From the definition of tensor product, it can be seen that Q in it's matrix form is given by $Q = Q_1 \otimes Q_2$. As argued in section 3, both Q_1 and Q_2 have the largest singular value 1. Also, from section 2, we see that $\sigma_{ij}(Q) = \sigma_i(Q_1)\sigma_j(Q_2)$. Using these observations, we can prove that the second largest singular value satisfies $\sigma_2(Q) = \max(\sigma_2(Q_1), \sigma_2(Q_2))$. Since the Renyi Correlation is the 2nd largest singular value of Q, this proves the main result of the paper.