Weighted Projection Quantiles Algorithm

November 19, 2014

Algorithm to calculate weighted projection quantile along the vector $\mathbf{u} \in \mathcal{B}_p$, given a set of observations $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$:

1. ***Compute $Q_{proj}(\mathbf{u})$, the projection quantile along $\mathbf{u}$***

   - Project each $\mathbf{X}_i$ along $\mathbf{u}$ to obtain $X_{\mathbf{u}i} = \frac{\langle \mathbf{X}_i, \mathbf{u} \rangle}{\|\mathbf{u}\|}$, for $i = 1, 2, ..., n$.
   - Find $\alpha = \frac{1+\|\mathbf{u}\|}{2}$-th quantile of $X_{\mathbf{u}1}, ..., X_{\mathbf{u}n}$, say $q_{\mathbf{u}}$..
   - $Q_{proj}(\mathbf{u}) = q_{\mathbf{u}} \mathbf{e}_{\mathbf{u}}$, $\mathbf{e}_{\mathbf{u}} = \mathbf{u}/\|\mathbf{u}\|$ being the unit vector along $\mathbf{u}$.

2. ***Compute Weights corresponding to this projection quantile $Q_{proj}(\mathbf{u})$***

   - Compute global weights for the direction vector $\mathbf{u}$ by k-mean distance:
     - Compute k-mean distance corresponding to $Q_{proj}(\mathbf{u})$ using $\bar{d}_k = \frac{1}{n} \sum_{i=1}^{n} d_i \mathbb{I}_{\{d_i < d_{(k)}\}}$, where d_i is the euclidean distace of $\mathbf{X}_i$ from $Q_{proj}(\mathbf{u})$ given by $\|\mathbf{X}_i - Q_{proj}(\mathbf{u})\|$. k is a tuning parameter.

     - Compute the weights corresponding to $\mathbf{u}$:

       $$w_{\mathbf{u}} = \exp(-a.d_k)$$

       where a is a tuning parameter.
   - Compute weights for each sample point $\mathbf{X}_i; i = 1, 2, ..., n$:
     - Compute the orthogonal Norms by $\|\mathbf{X}_{\mathbf{u}\perp i}\| = \|\mathbf{X}_i - X_{\mathbf{u}i} \mathbf{e}_{\mathbf{u}}\|$.
     - Compute weight of i^{th} sample:

       $$w_{2i} = \exp\left[-b \frac{\|\mathbf{X}_{\mathbf{u}\perp i}\|}{\|\mathbf{X}_i\|}\right] \mathbb{I}_{\{\|\mathbf{X}_{\mathbf{u}\perp i}\| \leq \epsilon\}}$$

       b, ϵ being tuning parameters.

3. ***Compute the weighted projection quantile***

   - Suppose there are m observations with non-zero weights w_{2i}, with indices $i_1, i_2, ..., i_m$. Define $\tilde{X}_{\mathbf{u}i_j} = w_{\mathbf{u}} w_{2i_j} X_{\mathbf{u}i_j}$.
   - Find $\alpha = \frac{1+\|\mathbf{u}\|}{2}$-th quantile of $\tilde{X}_{\mathbf{u}i_1}, ..., \tilde{X}_{\mathbf{u}i_m}$. Let it be $\tilde{q}_{\mathbf{u}}$.
   - Find the weighted projection quantile as $\tilde{Q}_{proj}(\mathbf{u}) = \tilde{q}_{\mathbf{u}} \mathbf{e}_{\mathbf{u}}$.

Definition Given a random vector $\mathbf{X} \in \mathbb{R}^p$ that follows a multivariate distribution F, and a point $\mathbf{p} \in \mathbb{R}^p$, find $\alpha_{\mathbf{p}}$ such that $\|\mathbf{p}\|$ is the $\alpha_{\mathbf{p}}$-th quantile for the projection of $\mathbf{X}$ on $\mathbf{p}$, say $X_{\mathbf{p}}$. Then the **Projection Quantile Depth** (PQD) at $\mathbf{p}$ with respect to F is defined as

$$D(\mathbf{p}, F) = \exp(-\alpha_{\mathbf{p}})$$

Given data $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$, the PQD at a given $\mathbf{p}$ can be estimated by finding the two nearest points on either side of $\|\mathbf{p}\|$ along $\mathbf{p}$, say $\mathbf{p}_1, \mathbf{p}_2$, obtain their corresponding quantiles, say α_1, α_2 respectively, then estimate $\alpha_{\mathbf{p}}$ by a linear approximation:

$$\hat{\alpha}_{\mathbf{p}} = \frac{(\alpha_1 - \alpha_2)(\|\mathbf{p}\| - \|\mathbf{p}_1\|)}{\|\mathbf{p}_1\| - \|\mathbf{p}_2\|} + \alpha_1$$

and plugging it in the above definition.

Algorithm 1 Algorithm for PQD-based classification

procedure PQDClassifier(training data $\mathbf{X}_i \in \mathbb{R}^{n_i \times p}$ with class labels i; $i = 1, 2, ..., k$, new data $\mathbf{x}_{new} \in \mathbb{R}^p$)
- Set $i = 1$.
- *top*:
- Estimate from the sample the PQD of $\mathbf{p}$ with respect to the i^{th} population, say $D(\mathbf{x}_{new}, \mathbf{X}_i)$.
- **if** $i = k$ **then Stop**
- **else**
  - Set $i \leftarrow i + 1$, **goto** *top*
- Find c that maximizes the PQD of $\mathbf{x}_{new}$ w.r.t. all possible classes:

$$D(\mathbf{x}_{new}, \mathbf{X}_c) = \max\{D(\mathbf{x}_{new}, \mathbf{X}_i) : i = 1, 2, ..., k\}$$

- Assign class c to new data $\mathbf{x}_{new}$.

Note One can define a weighted version of PQD by replacing $X_{\mathbf{p}}$ by their weighted version $\tilde{X}_{\mathbf{p}}$. A weighted classification scheme follows similarly.

New notion of data depth

We define a data depth based on projection quantiles (which, of course, can be extended to the weighted versions), on the lines of Zuo's Projection Depth [ref?].

Definition Given a random vector $X \in \mathbb{R}^p$ that follows a multivariate distribution F, and a point $x \in \mathbb{R}^p$, define an outlyingness function

$$O(x,F) = \sup_{\|u\|=1} \left| F_u(x_u) - \frac{1}{2} \right|$$

where $F_{\mathbf{u}}$ is the distribution function of $X_{\mathbf{u}}$, and $x_u = \langle x, u \rangle / \|u\|$. Then we define the **Projection Quantile Depth** at x with respect to F as

$$PQD(x,F) = \frac{1}{1 + O(x,F)}$$

We now follow the order in Zuo's paper to obtain the properties of the PQD.

(P1) ***Affine invariance*** Say G is the cdf of $AX + b$. Then we have $O(Ax + b, G) = O(x, F)$ because for all $u \in \mathbb{R}^p$ we have

$$F(x) = G(Ax + b) \Rightarrow F_u(x_u) = G_{Au+b}\left((Ax+b)_{Au+b}\right)$$

where subscripting denotes projection.

(P2) ***Quasi-concavity*** for $0 < \lambda < 1, z = \lambda x + (1 - \lambda) y$,

$$\begin{aligned} F_u(z_u) &= F_{\mathbf{u}}(\lambda x_u + (1-\lambda) y_u) \\ \Rightarrow \left| F_u(z_u) - \frac{1}{2} \right| &\leq \begin{cases} \left|\max\{F_u(x_u), F_u(y_u)\} - \frac{1}{2}\right| & \text{if } F_u(z_u) < 1/2 \\ \left|\min\{F_u(x_u), F_u(y_u)\} - \frac{1}{2}\right| & \text{otherwise} \end{cases} \\ &\leq \max\left\{ \left| F_u(x_u) - \frac{1}{2} \right|, \left| F_u(y_u) - \frac{1}{2} \right| \right\} \end{aligned}$$

and we finally get $PQD(z,F) \geq \max\{PQD(x,F), PQD(y,F)\}$.

(P3) ***Vanishing at infinity*** For $\|x\| \to \infty$ we have $F_u(x_u) \to 1$ hence $PQD(x,F) \to 2/3$. So we can define $PQD^*(x,F) = PQD(x,F) - 2/3$ to obtain P3.

(P4) ***Maximized at center of symmetric*** F Halfspace symmetry of F w.r.t. center θ implies (halfspace) symmetry of F_u around θ_u for any $u \in \mathbb{R}^p$, so that $F_u(\theta_u) = 1/2$ and subsequently PQD achieves its max value 1 at θ.

Example 1 For $X \sim \mathcal{N}_2((0,0)', I_2)$ we have $X_u \sim N(0,1)$ for any $u \in \mathbb{R}^p : \|u\| = 1$. Due to the symmetry of the distribution function maximizing $|F_u(x_u) - 1/2|$ is equivalent to maximizing $F_u(x_u) = \Phi(x_u) = \Phi(\langle x, u \rangle)$.

Converting u to polar coordinates we can maximize over the angle θ that u makes with the X-axis, and can easily find that the above function maximizes at $\theta = \tan^{-1}(x_2/x_1)$, which is the direction for x itself. Thus for given x the sup is obtained at $u = x$, and we have

$$PQD(x, F) = \frac{1}{1 + \Phi(\|x\|) - 1/2} = \frac{2}{1 + 2\Phi(\|x\|)} \tag{1}$$

Fig. 1 shows a comparison of PQD with Zuo's PD for standard bivariate normal.

Note In place of standard bivariate normal, one can use any circularly symmetric F, i.e. a distribution that has the same marginal distribution F_0 along all one-dimensional projections to come up with an exact formula for $PQD(x, F)$. In that general scenario, $\Phi(.)$ will be replaced by $F_0(.)$ in (1).

A general class of projection based depth functions

Definition We consider a random variable $X \in \mathbb{R}^p$ following the distribution F. Define F_u as the projection of F towards any $u \in \mathbb{R}^p$. In this setup, we define a general class of outlyingness (or htped) function:

$$O(x, F) = \sup_{\|u\|=1} f(x_u, \theta_u) \tag{2}$$

where $f(.)$ is a function satisfying properties mentioned below. $x_u = \langle x, u \rangle / \|u\|$, and $\theta \in \Theta_u \subseteq \mathbb{R}^k$ is a set of parameters (or a single parameter) characterizing the distribution function F_u. Then we define the depth at x with respect to F as $D(x, F) = 1/(1 + O(x, F))$.
The necessary properties of $f(.)$ are formulated keeping necessary properties of a depth functions in mind [ref zuo-serf]:

(F1) ***Affine invariant*** $f(x_u, \theta_u) = f(Ax_u + b, \theta_u^{AX+b})$ for non-singular matrix $A \in \mathbb{R}^{p \times p}$. where subscripting denotes projection.

(F2) ***Quasi-convex in*** x_u for $0 < \lambda < 1, z = \lambda x + (1 - \lambda)y$,

$$f(z_u) = f(\lambda x_u + (1 - \lambda)y_u) \leq \max\{f(x_u, \theta_u), f(y_u, \theta_u)\}$$

from which quasi-concavity of $D(x, F)$ follows.

(F3) ***Monotonically increasing in*** x_u ???

(F4) ***Maximized at center of symmetric*** F ??

(F5) ***Uniformly continuous in both*** x_u ***and*** θ_u

Theorem 0 Given an outlyingness function like (1) and an $f(.)$ satisfying properties (F1)-(F5), the resulting depth function is:
(i) Affine invariant,
(ii) Quasi-concave, and as a result $D(x, F)$ decreases monotonically along any ray starting from the deepest point, [ref]
(iii) Maximized at the center of a symmetric F.
(iv) Approaches a limit c at infinity, i.e. $D(x, F) \to c$ as $\|x\| \to \infty$.
(v) Uniformly continuous.

The proofs easily follow from the corresponding properties of $f(.)$.

Now suppose that θ_{nu} is the estimate of θ_u from a size-n iid sample. Following Zuo, we state some properties regarding the convergence $\theta_{nu} \to \theta_u$.
(C0) $\sup_{\|u\|=1} \|\theta_u\| < \infty$ for $i = 1, 2, ..., k$.
(C1) $\sup_{\|u\|=1} \|\theta_{nu} - \theta_u\| = o_P(1)$
(C2) $\sup_{\|u\|=1} \|\theta_{nu} - \theta_u\| = o(1)$
(C3) $\sup_{\|u\|=1} \|\theta_{nu} - \theta_u\| = O_P(1/\sqrt{n})$

Again following the lines of Zuo's paper, we say that:

Theorem 1 Given the condition (C0) above (?):
(i) $\sup_{x \in \mathbb{R}^d} |D(x, F_n) - D(x, F)| = o_P(1)$ given (C2) holds.
(ii) $\sup_{x \in \mathbb{R}^d} |D(x, F_n) - D(x, F)| = o(1)$ given (C3) holds.
(iii) $\sup_{x \in \mathbb{R}^d} |D(x, F_n) - D(x, F)| = O(1/\sqrt{n})$ given (C4) holds.

Proof **(ii)** Consider a sequence of cdfs $\{F_n\}$ converging in distribution to the true CDF F. Then we have for $x \in \mathbb{R}^p$

$$|D(x, F_n) - D(x, F)| = \frac{|O(x, F_n) - O(x, F)|}{(1 + O(x, F_n))(1 + O(x, F))} \leq |O(x, F_n) - O(x, F)|$$

Now we have that

$$\begin{aligned} |O(x, F_n) - O(x, F)| &= \left| \sup_{\|u\|=1} f(x_u, \theta_{nu}) - \sup_{\|u\|=1} f(x_u, \theta_u) \right| \\ &\leq \sup_{\|u\|=1} |f(x_u, \theta_{nu}) - f(x_u, \theta_u)| \end{aligned}$$

Since $f(x_u, .)$ is uniformly continuous in θ_u and (C2) holds by assumption, it follows that for fixed x we have $O(x, F_n) \to O(x, F)$ as $n \to \infty$. This allows us to conclude that for any $M > 0$ and x such that $\|x\| \leq M$,

$$\sup_{\|x\| \leq M} |D(x, F_n) - D(x, F)| \to 0 \text{ as } n \to \infty$$

For $\|x\| > M$, we have $D(x, F) \to c$ as $\|x\| \to \infty$ by Part (iv) of Theorem 0. Moreover, $O(x, F_n) \geq f(x_u, \theta_{nu})$ for all $u < \textit{incomplete} >$.