# Ranking - minimalizacja ryzyka wypukłego

Wojciech Rejchel

Wisła 2009

# Opis problemu

- $Z = (X, Y)$, $Z' = (X', Y')$ - niezależne o tym samym rozkładzie
- $X \in \mathcal{X} \subset \mathbb{R}^d$, $Y \in \mathbb{R}$

# Opis problemu

- $Z = (X, Y)$, $Z' = (X', Y')$ - niezależne o tym samym rozkładzie
- $X \in \mathcal{X} \subset \mathbb{R}^d$, $Y \in \mathbb{R}$
- $X, X'$ - obserwowane wektory cech
- $Y, Y'$ - nieznane etykiety

# Opis problemu

- $Z = (X, Y)$, $Z' = (X', Y')$ - niezależne o tym samym rozkładzie
- $X \in \mathcal{X} \subset \mathbb{R}^d$, $Y \in \mathbb{R}$
- $X, X'$ - obserwowane wektory cech
- $Y, Y'$ - nieznane etykiety
- $Z$ jest "lepszy" od $Z'$, jeśli $Y > Y'$

# Opis problemu

- Reguła rangująca $f : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$

# Opis problemu

- Reguła rangująca $f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$
- $$\text{Przewidujemy } y > y' \text{ , jeśli } f(x, x') > 0$$

# Opis problemu

- Reguła rangująca $f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$
- $$\text{Przewidujemy } y > y' \text{, jeśli } f(x, x') > 0$$
- Minimalizacja ryzyka

$$L(f) = P(\operatorname{sgn}(Y - Y') f(X, X') < 0) \rightsquigarrow \arg\min_{f \in \mathcal{F}} L(f)$$

# Opis problemu

- $Z_1 = (X_1, Y_1), \ldots, Z_n = (X_n, Y_n)$ - niezależne kopie $Z$

# Opis problemu

- $Z_1 = (X_1, Y_1), \ldots, Z_n = (X_n, Y_n)$ - niezależne kopie $Z$
- Minimalizacja ryzyka empirycznego

$$L_n(f) = \frac{1}{n(n-1)} \sum_{i \neq j} \mathbb{I}[\,\mathrm{sgn}(Y_i - Y_j)\, f(X_i, X_j) < 0]$$

$$\rightsquigarrow \arg\min_{f \in \mathcal{F}} L_n(f)$$

# Wypukłe ryzyko

- $\psi$ - wypukła funkcja straty

# Wypukłe ryzyko

- $\psi$ - wypukła funkcja straty
- $A(f) = \mathbf{E}\,\psi[\,\mathrm{sgn}(Y - Y')\,f(X, X')] \rightsquigarrow f^* = \arg\min_{f \in \mathcal{F}} A(f)$

# Wypukłe ryzyko

- $\psi$ - wypukła funkcja straty
- $A(f) = \mathbf{E}\,\psi[\,\mathrm{sgn}(Y - Y')\, f(X, X')] \rightsquigarrow f^* = \arg\min_{f \in \mathcal{F}} A(f)$
- Minimalizacja ryzyka empirycznego

$$A_n(f) = \frac{1}{n(n-1)} \sum_{i \neq j} \psi_f(Z_i, Z_j),$$

$$\text{gdzie} \;\; \psi_f(z, z') = \psi[\mathrm{sgn}(y - y')\, f(x, x')]$$

# Wypukłe ryzyko

- $\psi$ - wypukła funkcja straty
- $A(f) = \mathbf{E}\,\psi[\,\mathrm{sgn}(Y - Y')\, f(X, X')] \rightsquigarrow f^* = \arg\min_{f \in \mathcal{F}} A(f)$
- Minimalizacja ryzyka empirycznego

$$A_n(f) = \frac{1}{n(n-1)} \sum_{i \neq j} \psi_f(Z_i, Z_j),$$

$$\text{gdzie } \ \psi_f(z, z') = \psi[\mathrm{sgn}(y - y')\, f(x, x')]$$

- $\rightsquigarrow f_n = \arg\min_{f \in \mathcal{F}} A_n(f)$

# Ryzyko i ryzyko względne

- Dla dowolnego $0 < \alpha < 1$

$$P\left(A(f_n) \leqslant A_n(f_n) + \frac{C(\mathcal{F},\psi)\ \ln \alpha^{-1}}{n^{\beta}}\right) \geqslant 1 - \alpha$$

# Ryzyko i ryzyko względne

- Dla dowolnego $0 < \alpha < 1$

$$P\left(A(f_n) \leqslant A_n(f_n) + \frac{C(\mathcal{F},\psi)\ \ln \alpha^{-1}}{n^{\beta}}\right) \geqslant 1 - \alpha$$

-

$$P\left(A(f_n) \leqslant A(f^*) + \frac{C(\mathcal{F},\psi)\ \ln \alpha^{-1}}{n^{\beta}}\right) \geqslant 1 - \alpha$$

# Ryzyko i ryzyko względne

- Dla dowolnego $0 < \alpha < 1$
$$P\left(A(f_n) \leqslant A_n(f_n) + \frac{C(\mathcal{F},\psi)\ \ln \alpha^{-1}}{n^{\beta}}\right) \geqslant 1-\alpha$$
- $$P\left(A(f_n) \leqslant A(f^*) + \frac{C(\mathcal{F},\psi)\ \ln \alpha^{-1}}{n^{\beta}}\right) \geqslant 1-\alpha$$
- $\beta = 1/2$ - Clemencon(2008)

# Dekompozycja Hoeffdinga

- $$A(f) - A_n(f) = 2P_n[A(f) - P\,\psi_f] - U_n(h_f)$$

# Dekompozycja Hoeffdinga

- $$A(f) - A_n(f) = 2P_n[A(f) - P\,\psi_f] - U_n(h_f)$$

- $$\begin{aligned} P\,\psi_f(z) &= \mathbf{E}[\psi_f(Z, Z')|Z = z] \\ P_n(g) &= \frac{1}{n}\sum_{i=1}^{n} g(Z_i) \end{aligned}$$

# Dekompozycja Hoeffdinga

- $$A(f) - A_n(f) = 2P_n[A(f) - P\,\psi_f] - U_n(h_f)$$

- $$\begin{aligned} P\,\psi_f(z) &= \mathbf{E}[\psi_f(Z,Z')|Z=z] \\ P_n(g) &= \frac{1}{n}\sum_{i=1}^{n} g(Z_i) \end{aligned}$$

- $$\begin{aligned} h_f(Z_1,Z_2) &= \psi_f(Z_1,Z_2) - P\psi_f(Z_1) - P\psi_f(Z_2) + A(f) \\ U_n(h_f) &= \frac{1}{n(n-1)}\sum_{i\neq j} h_f(Z_i,Z_j) \end{aligned}$$

# Składnik empiryczny

1. Nierówność Tallagranda

# Składnik empiryczny

1. Nierówność Tallagranda
   - $\exists_{B>0} \ \forall_{f \in \mathcal{F}} \qquad Var[P\psi_f(Z)] \leqslant B\, A(f)$

# Składnik empiryczny

1. Nierówność Tallagranda
   - $\exists_{B>0} \ \forall_{f \in \mathcal{F}} \quad Var[P\psi_f(Z)] \leqslant B\, A(f)$
   - $\exists_{B>0} \ \forall_{f \in \mathcal{F}} \quad Var[P\psi_f(Z) - P\psi_{f^*}(Z)] \leqslant B\, [A(f) - A(f^*)]$

# Składnik empiryczny

1. Nierówność Tallagranda
   - $\exists_{B>0} \ \forall_{f \in \mathcal{F}} \qquad Var[P\psi_f(Z)] \leqslant B\,A(f)$
   - $\exists_{B>0} \ \forall_{f \in \mathcal{F}} \qquad Var[P\psi_f(Z) - P\psi_{f^*}(Z)] \leqslant B\,[A(f) - A(f^*)]$
2. Ograniczenie punktów stałych funkcji podpierwiastkowych

# Składnik empiryczny

**Twierdzenie 1**

Dla dowolnych $K > 1$ oraz $0 < \alpha < 1$

$$P\left(A(f_n) \leqslant \frac{K}{K-1} P_n(P\psi_{f_n}) + D\,\frac{\ln n + \ln \alpha^{-1}}{n}\right) \geqslant 1 - \alpha.$$

# Zdegenerowany $U$-proces

- $$\left\{ U_n(h_f) = \frac{1}{n(n-1)} \sum_{i \neq j} h_f(Z_i, Z_j) \ : \ f \in \mathcal{F} \right\}$$

# Zdegenerowany $U$-proces

- $$\left\{ U_n(h_f) = \frac{1}{n(n-1)} \sum_{i \neq j} h_f(Z_i, Z_j) \;:\; f \in \mathcal{F} \right\}$$
- $$h_f(Z_i, Z_j) = \psi_f(Z_i, Z_j) - P\psi_f(Z_i) - P\psi_f(Z_j) + A(f).$$

## Zdegenerowany $U$-proces

- $$\left\{ U_n(h_f) = \frac{1}{n(n-1)} \sum_{i \neq j} h_f(Z_i, Z_j) \ : \ f \in \mathcal{F} \right\}$$

- $$h_f(Z_i, Z_j) = \psi_f(Z_i, Z_j) - P\psi_f(Z_i) - P\psi_f(Z_j) + A(f).$$

- $$\mathbf{E} \left( \sup_{f \in \mathcal{F}} |U_n(h_f)| \right)^p \leqslant C \, \mathbf{E}_Z \, \mathbf{E}_\sigma \left( \sup_{f \in \mathcal{F}} |S_n(h_f)| \right)^p , \quad p \geqslant 1$$

# Zdegenerowany $U$-proces

- $$\left\{ U_n(h_f) = \frac{1}{n(n-1)} \sum_{i \neq j} h_f(Z_i, Z_j) \ : \ f \in \mathcal{F} \right\}$$

- $$h_f(Z_i, Z_j) = \psi_f(Z_i, Z_j) - P\psi_f(Z_i) - P\psi_f(Z_j) + A(f).$$

- $$\mathbf{E} \left( \sup_{f \in \mathcal{F}} |U_n(h_f)| \right)^p \leqslant C\, \mathbf{E}_Z\, \mathbf{E}_\sigma \left( \sup_{f \in \mathcal{F}} |S_n(h_f)| \right)^p, \quad p \geqslant 1$$

- $$S_n(h_f) = \frac{1}{n(n-1)} \sum_{i \neq j} \sigma_i \sigma_j h_f(Z_i, Z_j)$$

# Zdegenerowany $U$-proces

**Twierdzenie 2**

Dla dowolnej $0 < \alpha < 1$

$$P\left(U_n(h_{f_n}) \leqslant D\,\frac{\ln \alpha^{-1}}{n}\right) \geqslant 1 - \alpha.$$

# Nierówność probabilistyczna dla ryzyka

**Twierdzenie 3**

Dla dowolnej $0 < \alpha < 1$

$$P\left(A(f_n) \leqslant C\, A_n(f_n) + D\,\frac{\ln n + \ln \alpha^{-1}}{n}\right) \geqslant 1 - \alpha.$$

# Nierówność probabilistyczna dla względnego ryzyka

**Twierdzenie 4**

$\mathcal{F}$ - wypukła oraz $\psi$ - "ściśle wypukła". Dla dowolnej $0 < \alpha < 1$

$$P\left(A(f_n) \leqslant A(f^*) + D\,\frac{\ln n + \ln \alpha^{-1}}{n}\right) \geqslant 1 - \alpha.$$

# Przykłady

- Liniowe reguły rangujące

$$\mathcal{F} = \{f(x, x') = a^T(x - x') + b : a, x, x' \in \mathbb{R}^d, b \in \mathbb{R}\}$$

# Przykłady

- Liniowe reguły rangujące

$$\mathcal{F} = \{f(x,x') = a^T(x - x') + b : a, x, x' \in \mathbb{R}^d, b \in \mathbb{R}\}$$

- AdaBoost

# Przykłady

- Liniowe reguły rangujące

$$\mathcal{F} = \{f(x, x') = a^T(x - x') + b : a, x, x' \in \mathbb{R}^d, b \in \mathbb{R}\}$$

- AdaBoost

  $\psi(x) = \exp(-x)$

# Przykłady

- Liniowe reguły rangujące

$$\mathcal{F} = \{f(x, x') = a^T(x - x') + b : a, x, x' \in \mathbb{R}^d, b \in \mathbb{R}\}$$

- AdaBoost

  $\psi(x) = \exp(-x)$

  $\mathcal{R}$ - skończony VC-wymiar

# Przykłady

- Liniowe reguły rangujące

$$\mathcal{F} = \{f(x,x') = a^T(x - x') + b : a, x, x' \in \mathbb{R}^d, b \in \mathbb{R}\}$$

- AdaBoost

  $\psi(x) = \exp(-x)$

  $\mathcal{R}$ - skończony VC-wymiar

-

$$conv_T(\mathcal{R}) = \{f(x,x') = \sum_{j=1}^{T} w_j r_j(x,x') : \sum_{j=1}^{T} w_j = 1,$$
$$0 \leqslant w_j \leqslant 1, r_j \in \mathcal{R} \text{ dla } j = 1, \ldots, T\}$$

# Pytania

- Czy reguła $f_n$ jest także "dobra" dla $0-1$ funkcji straty?

# Pytania

- Czy reguła $f_n$ jest także "dobra" dla $0-1$ funkcji straty?
- $L(f) \leqslant A(f)$

# Bibliografia

M. A. Arcones, E. Gine, *U-processes indexed by Vapnik-Chervonenkis classes of functions with applications to asymptotics and bootstrap of U-statistics with estimated parameters*, Stochastic Process. Appl., vol. 52, pp. 17-38, 1994.

P. L. Bartlett, O. Bousquet, S. Mendelson, *Local Rademacher complexities*, Ann. Statist., vol. 33, pp. 1497-1537, 2005.

P. L. Bartlett, M. I. Jordan, J. D. McAuliffe, *Convexity, Classification, and Risk Bounds,* Journal of the American Statistical Association, vol. 101, pp. 138-156, 2006.

S. Clemençon, G. Lugosi, N. Vayatis, *Ranking and empirical minimization of U-statistics*, Ann. Statist., vol. 36, pp. 844-874, 2008.

Y. Freund, R. Iyer, R. E. Schapire, Y. Singer, *An efficient boosting algorithm for combining preferences*, J. Machine Learning Research, vol. 4, pp. 933-969, 2004.

V. H. de la Pena, E. Gine, *Decoupling: from dependence to independence.* Springer-Verlag, New York, 1999.