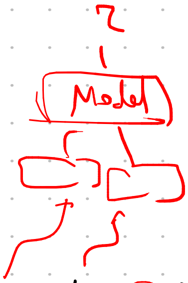
1

query → Google → $d_1, d_2, \dots, d_N$

$d_1 \to z_1$
$d_2 \to z_2$
$\vdots$
$d_k \to z_k$

$f(q, d_i) = z$

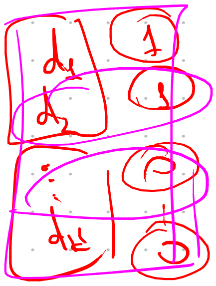


$D = \{(q_n, d_n, z_n)\}_{n=1}^{N}$ – regression

pointwise error function

q – d_1 — d_k

classification

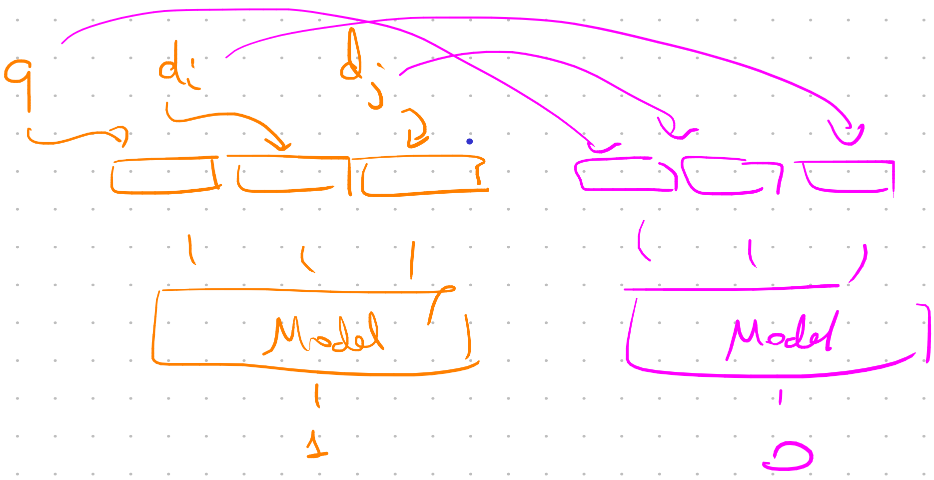


Learning to rank

pairwise (q, d_i, d_j) $\quad d_i \succ_q d_j$

listwise $(q, d_1, d_2, \dots, d_k)$

groupwise $(q, D_i \subset D)$



RankBoost

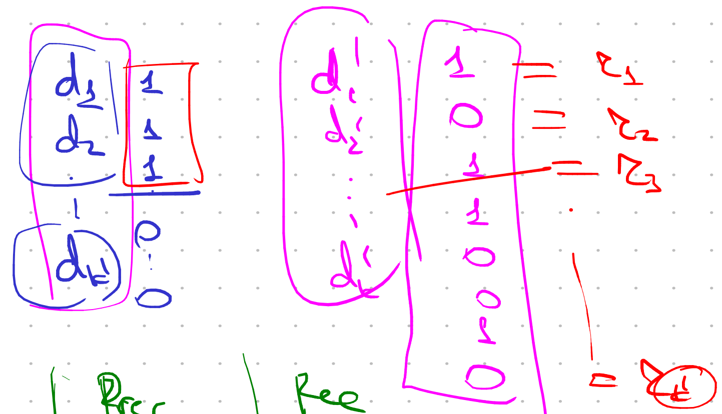
RankSVM

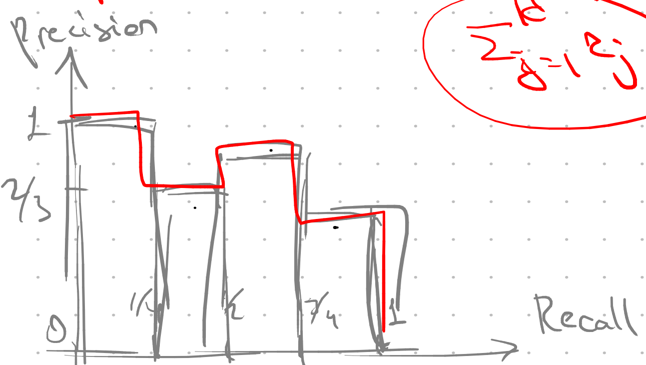
2 Quality metrics

$D - (q, d_1, d_2, \dots, d_k)$

$$\text{Precision@}k = \frac{\sum_{i=1}^{k} z_i}{k}$$

$$\text{Recall@}k = \frac{\sum_{i=1}^{k} z_i}{\sum_{j=1}^{k} z_j}$$





		Prec	Rec
d	1	1	1/4
	0	1/2	1/4
	1	2/3	1/2
	1	3/4	3/4
	0	3/5	3/4
	0	1/2	3/4
	1	4/7	1
d_8	0	1/2	1

Object detection

$$AP@K = \sum_{i=1}^{m} \text{Prec}_{[\text{Recall} = \frac{i}{m}]} \cdot \frac{1}{m} = \sum_{i=1}^{K} \text{Prec}@i \cdot \frac{[r_i = 1]}{m}$$

AUC = „area under curve"

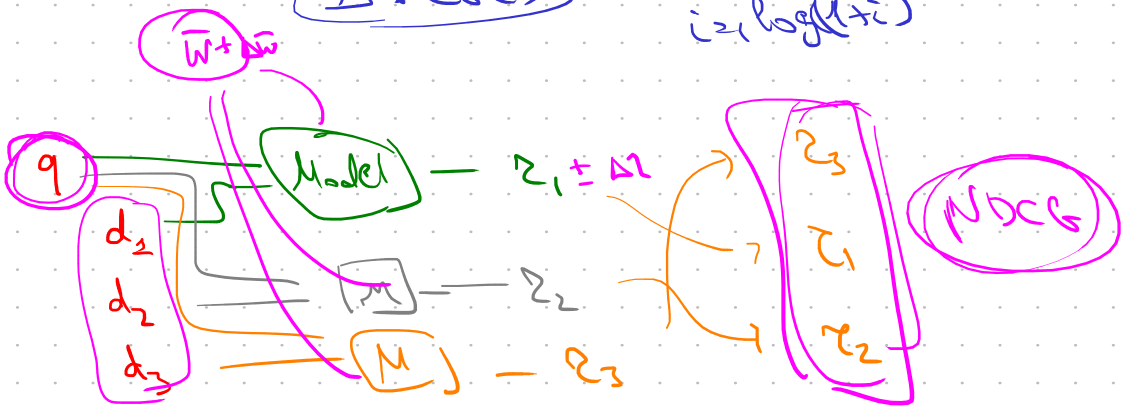
$$= \Pr[\text{2 случ. док-та будут отранж. прав}] = \frac{\#\{(d_i, d_j) : d_i > d_j,\ i < j\}}{\#\{(d_i, d_j) : r_i \neq r_j\}}$$

DCG – discounted cumulative gain

$$DCG(k) = \sum_{i=1}^{k} \frac{[r_i = 1]}{\log(1+i)}$$

normalized
↓

$$NDCG(k) = \frac{DCG(k)}{IDCG(k)}, \quad IDCG(k) = \sum_{i=1}^{\min(k, m)} \frac{1}{\log(1+i)}$$

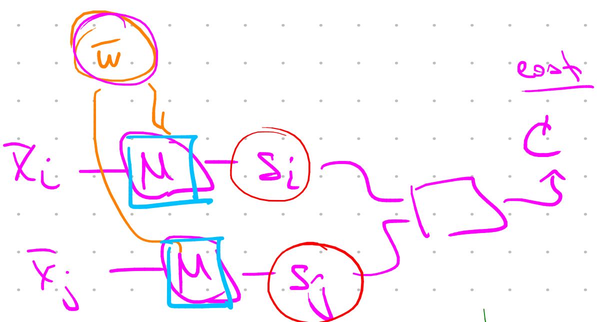


RankNet → LambdaRank → LambdaMART

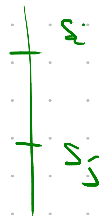
③ RankNet (pairwise)

$$\bar{x}_i \rightsquigarrow s_i = f(\bar{x}_i)$$

$$\bar{x}_j \rightsquigarrow s_j = f(\bar{x}_j)$$



$$\frac{\partial C}{\partial w_k} = \frac{\partial C}{\partial s_i} \frac{\partial s_i}{\partial w_k} + \frac{\partial C}{\partial s_j} \frac{\partial s_j}{\partial w_k}$$



$$p_{ij} = p(\bar{x}_i > \bar{x}_j) = \frac{1}{1 + e^{-\alpha(s_i - s_j)}} = \sigma(\alpha(s_i - s_j))$$

$$\boxed{r_{ij} = [\bar{x}_i > \bar{x}_j]} = 1$$

$(i,j) \to (i,j,1)$

$(j,i,0)$

r_{ij}

$$C = -r_{ij}\log p_{ij} - (1-r_{ij})\log(1-p_{ij}) =$$

$$= -r_{ij}\log\frac{1}{(1+e^{-\alpha(s_i-s_j)})} - (1-r_{ij})\log\frac{e^{-\alpha(s_i-s_j)}}{(1+e^{-\alpha(s_i-s_j)})}$$

$$= \log(1+e^{-\alpha(s_i-s_j)}) + (1-r_{ij})\cdot\alpha(s_i-s_j)$$

$$= [\text{bce } r_{ij}=1] = \log(1+e^{-\alpha(s_i-s_j)})$$

$$\frac{\partial C}{\partial s_i} = \frac{-\alpha\cdot e^{-\alpha(s_i-s_j)}}{1+e^{-\alpha(s_i-s_j)}} = -\frac{\alpha}{1+e^{\alpha(s_i-s_j)}}$$

$$\frac{\partial C}{\partial s_j} = \frac{+\alpha\cdot e^{-\alpha(s_i-s_j)}}{1+e^{-\alpha(s_i-s_j)}} = -\frac{\partial C}{\partial s_i}$$

$$\frac{\partial C}{\partial w_k} = \boxed{\frac{\partial C}{\partial s_i}}\frac{\partial s_i}{\partial w_k} + \boxed{\frac{\partial C}{\partial s_j}}\cdot\frac{\partial s_j}{\partial w_k} =$$

$$= -\boxed{\frac{\alpha}{1+e^{\alpha(s_i-s_j)}}}\left(\frac{\partial s_i}{\partial w_k} - \frac{\partial s_j}{\partial w_k}\right)$$

$$w_k = w_k - \eta\cdot\partial C/\partial w_k \quad \text{"} \quad \lambda_{ij}$$

$$\Delta w_k = \eta\cdot\sum_{(i,j)\in D}\lambda_{ij}\left(\frac{\partial s_i}{\partial w_k} - \frac{\partial s_j}{\partial w_k}\right) =$$

$$\boxed{\Delta w_k = \eta\cdot\sum_{i=1}^{n}\lambda_i\cdot\frac{\partial s_i}{\partial w_k}},$$

$$\lambda_i = \left(\sum_{(i,j)\in D}\lambda_{ij} - \sum_{(j,i)\in D}\lambda_{ji}\right)$$

$$C(i,j) = \log(1+e^{-\alpha(s_i-s_j)})$$

$$C(j,i) = \log(1+e^{-\alpha(s_j-s_i)}) + \alpha(s_j-s_i)$$

$$= -\log\frac{1}{1+e^{-\alpha(s_j-s_i)}} - \log e^{\alpha(s_j-s_i)}$$

$$= -\log\frac{e^{\alpha(s_j-s_i)}}{1+e^{-\alpha(s_j-s_i)}} =$$

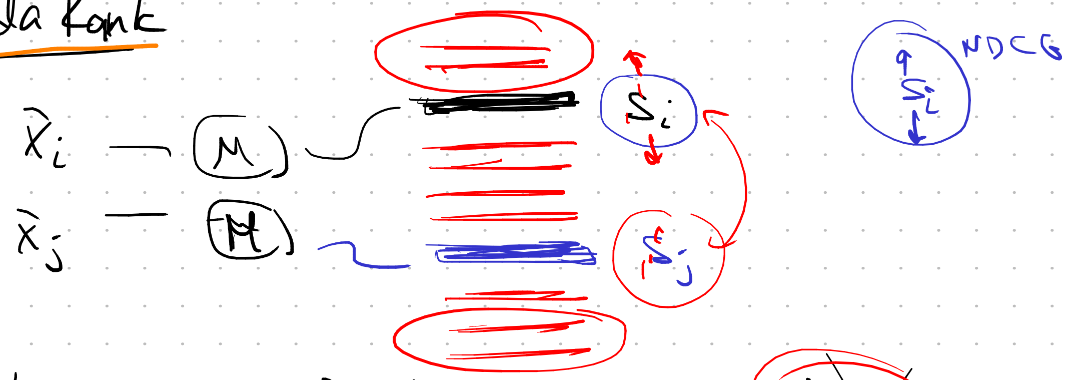
$$= -\log\frac{1}{1+e^{\alpha(s_j-s_i)}}$$

1) $\equiv$ $n \times f(\bar{x})$, $s_1 \dots s_n$

2) n^2, λ_{ij}, λ_i

3) n: $\partial s_i/\partial w_k$

4 LambdaRank



$$\lambda_{ij} = \frac{\alpha}{1+e^{\alpha(s_i-s_j)}} = \frac{\partial c(s_i-s_j)}{\partial s_i} \;\; \cancel{\times} \;\; \cancel{\frac{\partial NDCG}{\partial s_i}}$$

$$\lambda'_{ij} = \lambda_{ij} \cdot |\Delta NDCG(i,j)|$$

$$C(\bar{w}) \xrightarrow{\bar{w}} \min \qquad \frac{\partial C}{\partial w_k} = \sum_{i=1}^{n} \lambda_i \frac{\partial s_i}{\partial w_k} =$$

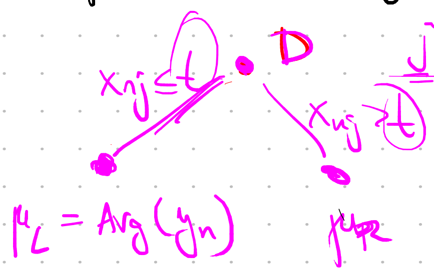
$$= \sum_{i=1}^{n} \left(\sum_{i,j \in D} \lambda_{ij} - \sum_{j,i \in D} \lambda_{ji} \right) \cdot \frac{\partial s_i}{\partial w_k}$$

$$\frac{\partial C'}{\partial w_k} = \sum_i \underbrace{\left(\sum \lambda'_{ij} - \sum \lambda'_{ji} \right)}_{\partial / \partial w_e}^{\lambda'_i} \boxed{\frac{\partial s_i}{\partial w_k}} \qquad \lambda'_{ij} \quad \lambda'_{ij}$$

$$\frac{\partial F}{\partial x_i} = f_i, \text{ если } \frac{\partial f_i}{\partial x_j} = \frac{\partial f_j}{\partial x_i} \;\; \forall_{ij}$$

5 MART — multiple additive regression trees

$$D = \{(\bar{x}_n, y_n)\}_{n=1}^{N}$$



$j = 1..d, \; \bar{x}_n \in \mathbb{R}^d, \; t$

$$L_j = \sum_{n \in L} (y_n - \mu_L)^2 + \sum_{n \in R} (y_n - \mu_R)^2 \xrightarrow[t_j]{} \min$$

L листьев

$$F_N(\bar{x}) = \sum_{i=1}^{N} \alpha_i f_i(\bar{x})$$

$$F_n(\bar{x}) \rightsquigarrow F_{n+1}(\bar{x}) = F_n(\bar{x}) + \alpha_{n+1} f_{n+1}(\bar{x})$$

C – ф-ция ошибки

$$\underline{C(F_n(\bar{x}), y_n)}$$

$\forall \bar{x}_i \quad f_{n+1}(\bar{x}_i) = F_n(\bar{x}_i) + \eta \Delta f(\bar{x}_i)$

$$\underline{\Delta C(\bar{x}_i)} \approx \Delta f(\bar{x}_i) \cdot \boxed{\frac{\partial C}{\partial F_n}\Big|_{F_n(\bar{x}_i)}}$$

$$\underline{\Delta f(\bar{x}_i)} \approx \underline{-\eta \cdot \frac{\partial C}{\partial f_n}\Big|_{F_n(\bar{x}_i)}}$$

$$D_{n+1} = \left\{ \left(\bar{x}_i, -\frac{\partial C(F_n)}{\partial f_n}\Big|_{F_n(\bar{x}_i)} \right) \right\}_{i=1}^{M}$$

Обучаем f_{n+1} на D_{n+1}

$$F_{n+1}(\bar{x}) = F_n(\bar{x}) + \alpha_{n+1} f_{n+1}(\bar{x})$$

$(\bar{x}_i, y_i), \quad y_i \in \{\pm 1\}$

$$p(y=1 \mid \bar{x}) = \boxed{\frac{1}{1+e^{-2\alpha F_N(\bar{x})}}} = p_+$$

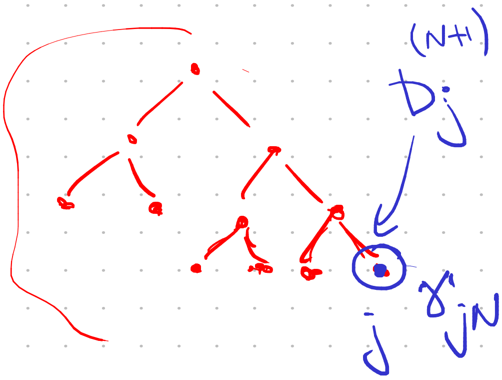
$$p_- = p(y=0 \mid \bar{x}) = 1 - p_+ = \frac{e^{-2\alpha F_N(\bar{x})}}{1+e^{-2\alpha F_N(\bar{x})}} = \frac{1}{1+e^{+2\alpha F_N(\bar{x})}}$$

$$L(F, y_i) = -\underbrace{[y_i=1]}\cdot \log p_+ - \underbrace{[y_i=-1]}\cdot \log p_- =$$

$$= -\log \frac{1}{1+e^{-2\alpha \cdot y_i F_N(\bar{x}_i)}} = \log\left(1+e^{-2\alpha y_i F_N(\bar{x}_i)}\right)$$

$$y_i^{(N+1)} = -\frac{\partial L}{\partial F}\Big|_{F_N(\bar{x}_i)} = \frac{2\alpha y_i \cdot e^{-2\alpha y_i F_N(\bar{x}_i)}}{1+e^{-2\alpha y_i F_N(\bar{x}_i)}} = \frac{2\alpha y_i}{1+e^{\beta_i}}$$

$f_{N+1}(\bar{x})$ обучен на $\underline{D^{(N+1)} = \{(\bar{x}_i, y_i^{(N+1)})\}}_{i=1}^{M}$



$$\gamma_{jN} = \underset{\gamma}{\operatorname{argmin}} \sum_{\underline{\bar{x}_i \in D_j^{(N+1)}}} \log\left(1+e^{-2\alpha y_i(F_N(\bar{x}_i)+\gamma)}\right)$$

$$= \underset{\gamma}{\operatorname{argmin}}\ g(\gamma)$$

$$\boxed{\gamma := -\frac{g'(\gamma)}{g''(\gamma)}}$$ Newton

$$-\sum_i \frac{2\alpha y_i}{1+e^{\beta_i}} = -\sum_i y_i^{(N+1)}$$

$$\parallel$$

$$g'(\gamma) = \sum_i \frac{1}{1+e^{\dots}} \cdot \left(-2\alpha y_i \cdot e^{-2\alpha\dots}\right) = \sum_i \frac{-2\alpha y_i}{1+e^{2\alpha y_i(F_i+\gamma)}}$$

$$g''(\gamma) = \sum_i \frac{+2\alpha y_i}{\left(1+e^{2\alpha y_i(F_i+\gamma)}\right)^2} \cdot 2\alpha y_i \cdot e^{\overbrace{2\alpha y_i(F_i+\gamma)}^{\beta_i}} = \sum_i \frac{4\alpha^2 y_i^2 \cdot e^{\beta_i}}{(1+e^{\beta_i})^2}$$

$$|y_i^{(N+1)}| \cdot \frac{2\alpha y_i e^{\beta_i}}{1+e^{\beta_i}} = |y_i^{(N+1)}|\left(2\alpha - |y_i^{(N+1)}|\right)$$

$$\gamma := \frac{\cancel{2\alpha} \sum_i y_i / \cancel{1+e^{\beta_i}}}{4\alpha^2 \sum_i \frac{y_i^2 e^{\beta_i}}{(1+e^{\beta_i})^2}} \quad \Leftarrow$$

$$\boxed{\gamma = \frac{\sum_i y_i^{(N+1)}}{\sum_i |y_i^{(N+1)}|\left(2\alpha - |y_i^{(N+1)}|\right)}}$$

5 Lambda MART

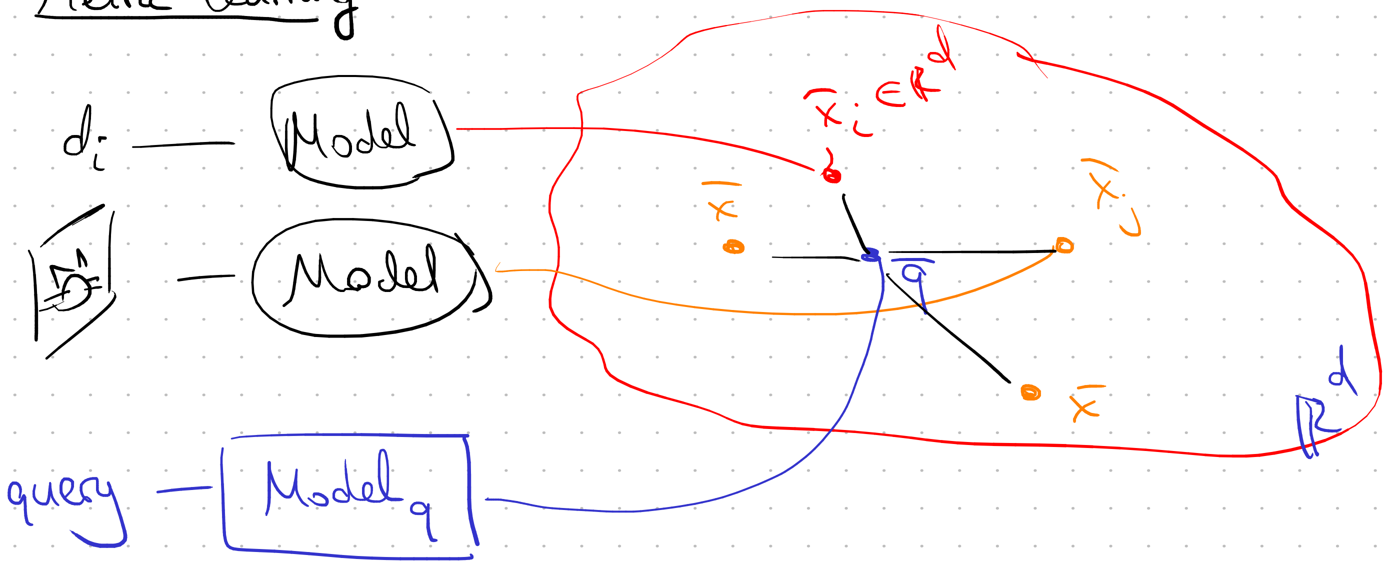
$$y_i^{(N+1)} = \frac{\partial C}{\partial s_i} = \lambda_i$$ – MART for regression

$$y_i^{(N+1)} = \underline{\lambda_i'} = \left(\sum \lambda_{ij} \underline{|\Delta NDCG_{ij}|} - \sum \lambda_{ji} |\Delta ..|\right) = \frac{\partial C'}{\partial s_i}$$

LambdaMART

$$\gamma_{jN} = \frac{\sum_i \lambda_i'}{\sum_i |\lambda_i'|(2d - |\lambda_i'|)}$$

6 Metric learning



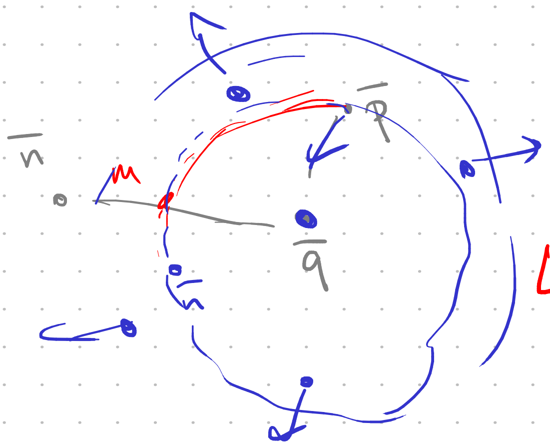
$(\bar{q}, \bar{x}_j, z)$: $z=1 \Rightarrow d(\bar{q}, \bar{x}) \to \min$

$z=0 \Rightarrow d(\bar{q}, \bar{x}) \to \max$

Triplet Loss $(\bar{q}, \bar{p}, \bar{n})$ – triplet

$\cancel{\underline{d(\bar{p}, \bar{q}) - d(\bar{n}, \bar{q})} \to \min}$

margin

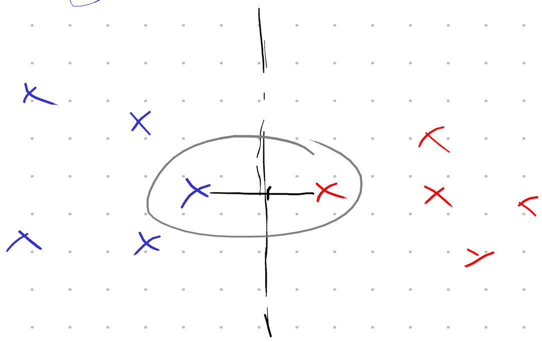


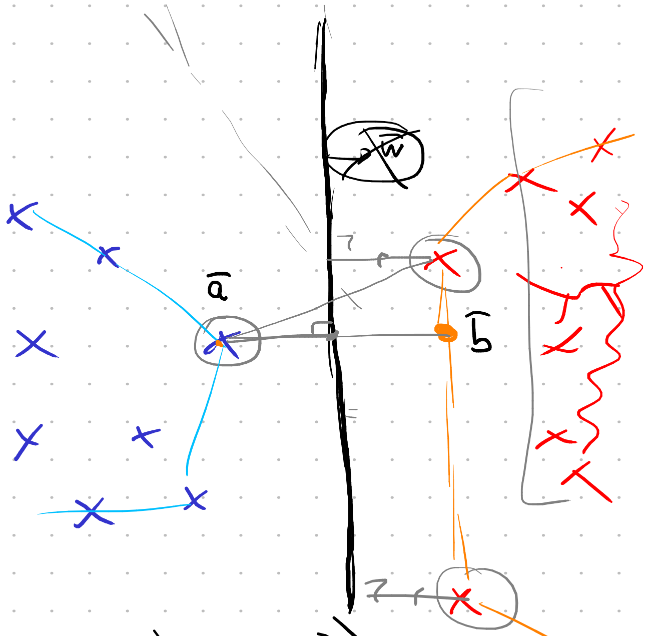
$$d(\bar{n}, \bar{q}) \geq d(\bar{p}, \bar{q}) + m$$ triplet loss

$$L = \max(0, d(\bar{p}, \bar{q}) - \underline{d(\bar{n}, \bar{q})} + m)$$

Hard negatives mining

① Support vector machines





$$D=\{(\bar x_n, t_n)\}$$

$$C_1=\{\bar x_n,\, t_n=1\}$$

$$C_2=\{\bar x_n,\, t_n=-1\}$$

$$\min \|\bar a-\bar b\|^2$$

$\bar a\in Conv(C_1)$

$\bar b\in Conv(C_2)$

$$\bar a=\sum_{n:t_n=1}\alpha_n\bar x_n, \quad \forall n\ \alpha_n\geq 0$$

$$\sum_{t_n=1}\alpha_n=1=\sum_{t_n=-1}\alpha_n$$

$$\bar b=\sum_{n:t_n=-1}\alpha_n\bar x_n,$$

$$\sum \alpha_n t_n \bar x_n$$

$$\|$$

quadratic programming

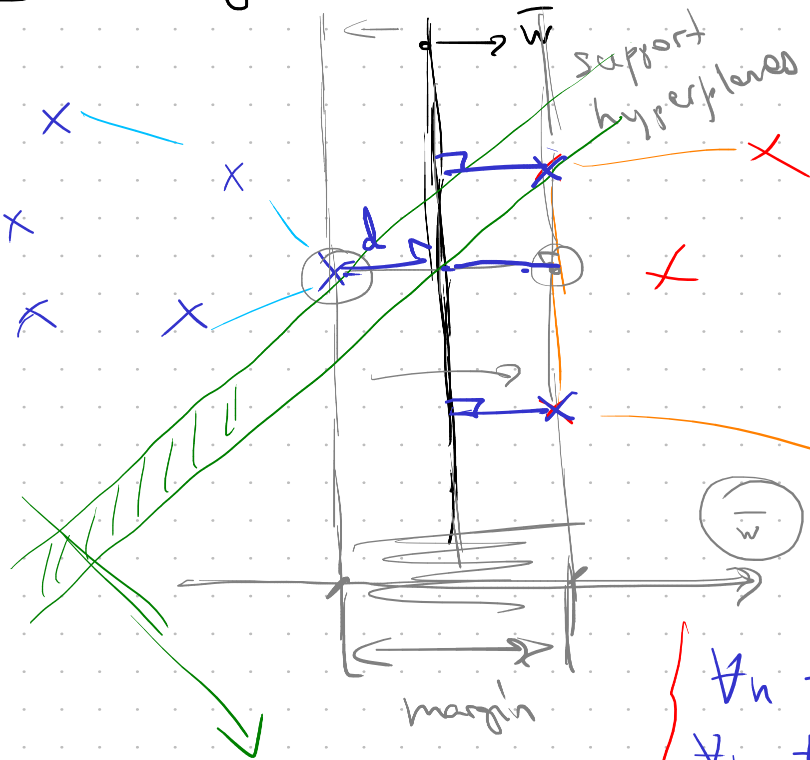
$$\min_{\bar\alpha}\left\|\sum_{t_n=1}\alpha_n\bar x_n-\sum_{t_n=-1}\alpha_n\bar x_n\right\|^2$$

N variables

при условии $\forall n\ \alpha_n\geq 0$

$$\sum_{t_n=1}\alpha_n=\sum_{t_n=-1}\alpha_n=1$$

② Margin minimization



A: $y = \bar w^T \bar x + w_0$

$$d(\bar x, A) = \frac{|\bar w^T \bar x + w_0|}{\|\bar w\|}$$

$$\bar w, w_0 = \arg\max_{\bar w, w_0} \min_n d(\bar x_n, A)$$

$$= \arg\max_{\bar w, w_0} \min_n \frac{|\bar w^T \bar x + w_0|}{\|\bar w\|}$$

при условии

$$\begin{cases} \forall n \quad t_n = 1 & \bar w^T \bar x_n + w_0 > 0 \\ \forall n \quad t_n = -1 & \bar w^T \bar x_n + w_0 < 0 \end{cases}$$

$$\boxed{\forall n \quad t_n(\bar w^T \bar x_n + w_0) > 0}$$

$$\arg\max_{\bar w, w_0} \min_n \frac{t_n(\bar w^T \bar x_n + w_0)}{\|\bar w\|} = \arg\max_{\bar w, w_0} \frac{1}{\|\bar w\|} \cdot \min_n t_n(\bar w^T \bar x_n + w_0)$$

$\|\bar w\| = 1$: $\cancel{\arg\max \min_n t_n(\bar w^T \bar x_n + w_0)}$

$\cancel{\forall n \quad t_n(\ldots) \geq 0}$

$\|\bar w\|^2 = 1$

fix $\|\bar w\|$:

$$\min_n t_n(\bar w^T \bar x_n + w_0) = 1$$

$$\arg\max_{\bar w, \bar w_0} \frac{1}{\|\bar w\|} = \arg\min_{\bar w} \|\bar w\|^2$$

при условии $\quad \forall n \quad t_n(\bar w^T \bar x_n + w_0) \geq 1$

quadratic programming

d variables

d^2, d^3, ∞