Crammer & Singer:

$Y = \{1, 2, 3, \dots\}$

EQ(4) $\forall 1 \le n \le N$ and $\forall$ classes $1 \le \bar{y} \le |Y|$ no. of classes

$$\forall n \; \forall \bar{y} \quad \boxed{M_{\bar{y}} \cdot x}^{w^T \phi(x_n, \bar{y})} + \delta_{y, \bar{y}} - \underbrace{M_y x}_{w^T \phi(x_n, y_n)} \ge 1$$

→ redundant for $\bar{y} = y$!

Rank SVM

EQ(10+11) $\forall (d_i, d_j) \in r_1^*: \quad w^T \phi(q_1, d_i) \ge w^T \phi(q_1, d_j)$

$\vdots$

$\forall (d_i, d_j) \in r_N^*: \quad w^T \phi(q_N, d_i) \ge w^T \phi(q_N, d_j)$

AUC SVM

EQ(4): $\forall n^+$ and $\forall n^-$: $\quad w^T \phi(x^+) - w^T \phi(x^-) > 0$

TRY 1:

any row of $M \rightarrow w_y$

$\arg\max_j \{M_j \cdot x\}$

$w_y \cdot x_{\in y} > w_{\bar{y}}\, x_{\notin y}$

$\underbrace{w_y (x_{\in y} - x_{\notin y}) > 1}_{\text{any two set of small class}}$

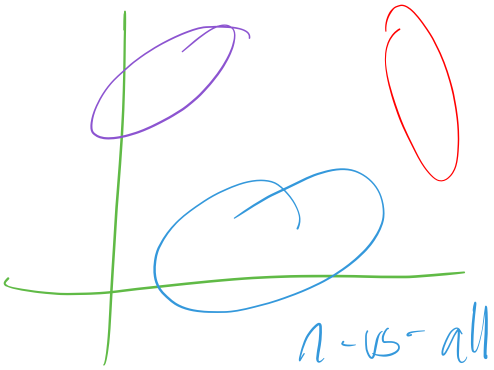
$w^T(\phi(q, d_i) - \phi(q, d_j))$

$w^T(\phi(x^+) - \phi(x^-)) > 1$

$\phi(x^+) = \phi(q_n, d_i)$

$\phi(x^-) = \phi(q_n, d_j)$

$\phi(q_n) \otimes \phi(d_i)$



AUC + Ranking ✓

$w^T \phi(x^+) > w^T \phi(x^-)$

Crammer & Singer:

$w = (m_{11}\, m_{12} \dots m_{1D},\, m_{21}\, m_{22}, \dots w_{2D}, \dots m_{|Y|1} \dots m_{|Y|D})^T$

$= (w_1 \dots \dots w_{|Y| \cdot D})^T$

class 1 $\qquad$ class 2 $\qquad$ class $|Y|$

$$\Phi(x, y) \xrightarrow{\phi} (\overset{1}{0}, \overset{2}{0}, \overset{3}{0}, \dots, 0, \overset{D+1}{x_1}, \overset{D+2}{x_2}, \dots, \overset{D+D}{x_D}, \overset{2D+1}{0}, 0, \dots, 0)^T$$

$y = 2$

Starting pos for class $\bar{y}$:

INPUT: $D(\bar{y}-1)+1$

y	x
2	$D+1: x_1 \quad D+2: x_2 \quad \dots$

Sparse encoding

only represent the non-negative

$$\to f(x, y) = w^T \phi(x, y)$$

$$\to \arg\max_{\bar{y} \in Y} f(x, \bar{y})$$

Before: $\quad f(x) = w^T x + b$

Now: $\quad f(x, y) = w^T \phi(x, y)$

Constraints:

$$\forall n \;\; \forall \bar{y} \neq y_n : \quad \overbrace{w^T \phi(x_n, y_n)}^{f(x_n, y_n)} > \overbrace{w^T \phi(x_n, \bar{y})}^{f(x_n, \bar{y})} + 1$$

$n = 1 \dots N$

$$w^T \left(\Phi(x_n, y_n) - \Phi(x_n, \bar{y}) \right) \geq 1$$

learning algorithm: Perceptron!

q...

$$f(x,y) = w^T \Phi(x,y) \qquad \mathcal{D} = \{(x_n, y_n)\}_{n=1\dots N}$$

application driven!! see above!

init $w \leftarrow 0$

loop epoch

  loop $1 \dots N$

    if $\arg\max_{\bar{y}} f(x_n, \bar{y}) \neq y_n$

    then update: $w \leftarrow w + \Phi(x_n, y_n) - \Phi(x_n, \bar{y})$

  end

until converge

$$
\begin{array}{rcrrrr}
 & \Phi(x_n, y_n) = & 1 & 1 & 0 & 0 \\
- & \Phi(x_n, \bar{y}) = & 0 & 1 & 0 & 1 \\
\hline
 & & 1 & 0 & 0 & -1
\end{array}
$$

$$w_0 = 0$$

$$w_1 = \Phi(x_{n_1}, y_{n_1}) - \Phi(x_{n_1}, \bar{y})$$

$$w_2 = \Phi(x_{n_1}, y_{n_1}) - \Phi(x_{n_1}, \bar{y}) + \Phi(x_{n_2}, y_{n_2}) - \Phi(x_{n_2}, \bar{y})$$

$$= \dots$$

⇒ incorporate a constr $\alpha_n(\bar{y}) = 0$

$$w = \sum_{n=1}^{N} \sum_{\bar{y} \neq y_n} \alpha_n(\bar{y}) \left(\Phi(x_n, y_n) - \Phi(x_n, \bar{y}) \right)$$

$$f(x, y) = \sum \sum \alpha_n(\bar{y}) \left(\Phi(x_n, y_n) - \Phi(x_n, \bar{y}) \right)^T \phi(x, y)$$

$$= \sum \sum \alpha_n(\bar{y}) \; K(x_n, y_n, x, y) - K(x_n, \bar{y}, x, y)$$

infeasible summation (why?)

→ implementation runs only on observed pairs with $\alpha > 0$

imagine loss function:

$$\min_{w} \; E(w) = \sum_{n=1}^{N} V(x_n, y_n, f)$$

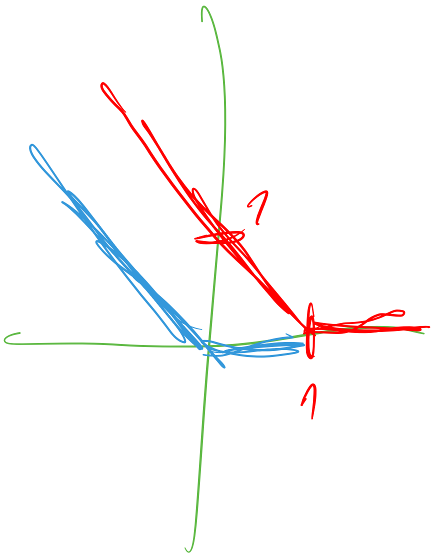
$$= \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}\left[\arg\max_{\bar{y}} f(x_n, \bar{y}) \neq y_n \right]$$

Vanilla perceptron:

if $\underline{y_n f(x_n)} < 0$ then update

compare hinge loss $(1 - y f(x))$

$$E(w) = \sum_{n=1}^{N} \max\{0, \overbrace{1 - y_n f(x_n)}\}$$



$$\frac{\partial E}{\partial w} = \begin{cases} 0 & \text{if } \dots \\ \underline{-y_n x} \end{cases} \Rightarrow \text{perceptron update}$$

$$w \leftarrow w + yx$$

$$\frac{\partial E}{\partial b} = \dots$$