Winnow Notes

Winnow Algorithm (Littlestone 1988)

Input: $\{(x_i, y_i)\}_i$, parameter $\eta > 0$

Initialize $w_j^{(0)} = \frac{1}{p}$ $j = 1 \ldots p$ (p features)

For $t = 1, 2, \ldots$

Locate i s.t. $y_i(\vec{w}^{(t)} \cdot \vec{x}_i) \leq 0$.

If none, stop and output $\vec{w}^{(t)}$

Else, $\forall j$, $w_j^{(t+1)} = \dfrac{w_j^{(t)} \cdot e^{\eta y_i x_{ij}}}{Z_t}$ ← normalization so $\sum_j w_j^{(t+1)} = 1$

End

idea: find a miscl point

reward features that agree with label: if $\operatorname{sign}(x_{ij}) = y_i$

$$w_j^{(t+1)} \propto w_j^{(t)} e^{(+)\text{ve}}$$

punish features that disagree: if - - - - - $\neq$.

$$w_j^{(t+1)} \propto w_j^{(t)} e^{(-)\text{ve}}$$

Winnow Convergence Bound:

Assume $\max_i \|x_i\|_\infty = 1$

Assume there is $\vec{w}^*$: $y_i(\vec{w}^* \cdot \vec{x}_i) \geq \delta \ \forall i$

and $\underbrace{w_j^* \geq 0 \ \forall j \text{ and } \|\vec{w}^*\|_1 = 1}_{\text{discrete prob distn}}$

Theorem: The Winnow alg makes at most

$$T \leq \frac{\ln p}{\eta\delta + \ln\left(\frac{2}{e^{\eta} + e^{-\eta}}\right)} \text{ mistakes.}$$

proof idea: show $\vec{w}^{(t)}$ is closer to $\vec{w}^*$ at each iteration,

in terms of $\underbrace{\text{KL divergence}}_{\text{"dist" between prob distns}}$.

$$KL(\vec{a} \| \vec{b}) := \sum_j a_j \log\left(\frac{a_j}{b_j}\right)$$

Why is it a "distance"? If $\vec{a} = \vec{b}$, $\log\left(\frac{a_j}{b_j}\right) = 0$ so $KL(\vec{a}, \vec{b}) = 0$

Turns out $KL(\vec{a}, \vec{b}) \geq 0$

$$\Phi_t = KL(w^* \| w^{(t)}) = \sum_j w_j^* \ln\left(\frac{w_j^*}{w_j^{(t)}}\right)$$

Step 1a
lower bound improvement

$$\Phi_t - \Phi_{t+1} = \sum_j w_j^* \ln\left(\frac{w_j^*}{w_j^{(t)}}\right) - \sum_j w_j^* \ln\left(\frac{w_j^*}{w_j^{(t+1)}}\right)$$

$$\overset{\text{simplify}}{=} \sum_j w_j^* \ln\left(\frac{w_j^{(t+1)}}{w_j^{(t)}}\right)$$

$$\overset{\text{alg}}{=} \sum_j w_j^* \ln\left(\frac{e^{\eta y_i x_{ij}}}{Z_t}\right)$$

$$= \sum_j w_j^* \underbrace{\ln e}^{\eta y_i x_{ij}} - \sum_j w_j^* \ln Z_t$$

$$= \sum_j w_j^* \eta y_i x_{ij} - \ln Z_t) \underbrace{\sum_j w_j^*}_{1}$$

$$= \underbrace{\eta y_i (w^* \cdot x_i)}_{\substack{\text{V1-def } w^* \\ \delta}} - \ln(\quad _t)$$

Go to aside for this step

$$\geq \eta \delta - \ln\left(\frac{e^{\eta} + e^{-\eta}}{2}\right) = \Lambda \quad \leftarrow \text{denom in bound}$$

Step 1b
add up improvements

$$\Phi_0 - \Phi_T = \sum_{t=0}^{T-1} (\Phi_t - \Phi_{t+1}) \geq T\Lambda$$

step 2

$$\Phi_0 - \Phi_T \underset{\substack{\uparrow \\ \text{KL div} \\ \text{is non neg}}}{\leq} \Phi_0 = K(w^* \| \bar{w}^0) = \sum_j w_j^* \ln\left(\frac{w_j^*}{1/p}\right) = \sum_j w_j^* \ln(p\, w_j^*)$$

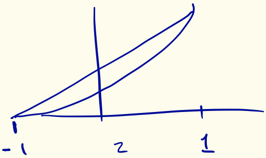
$$\underset{\substack{\uparrow \\ 0 \leq w_j^* \leq 1}}{\leq} \sum_j w_j^* \ln(p \cdot 1) = \underbrace{\sum_j w_j^*}_{1} \ln p = \ln p$$

So $\quad T\Lambda \leq \ln p \implies T \leq \frac{\ln p}{\Lambda} \quad \square$

aside:

$$1 = \sum_j w_j^{(t+1)} = \frac{\sum_j w_j^{(t)} \cdot e^{\eta y_i x_{ij}}}{Z_t}$$

$$Z_t = \sum_j w_j^{(t)} \cdot e^{\eta y_i x_{ij}}$$



Use a bound of exponential by a linear function

$$e^{\eta z} \le \left(\frac{1+z}{2}\right)e^{\eta} + \left(\frac{1-z}{2}\right)e^{-\eta}$$

holds for $-1 \le z \le 1$

$$Z_t \le \sum_j w_j^{(t)}\left(\frac{1+y_i x_{ij}}{2}\right)e^{\eta} + w_j^{(t)}\left(\frac{1-y_i x_{ij}}{2}\right)e^{-\eta}$$

$$\underset{\text{collect terms}}{=} \frac{e^{\eta}+e^{-\eta}}{2}\underbrace{\sum_j w_j^{(t)}}_{1} + \frac{e^{\eta}-e^{-\eta}}{2}\underbrace{\sum_j w_j^{(t)} y_i x_{ij}}_{\underbrace{y_i(\bar{x}_i \cdot \bar{w}^{(t)})}_{\le 0 \;-\; i \text{ is miscl at } t}}$$

$$\le \frac{e^{\eta}+e^{-\eta}}{2}$$

$$-\ln(Z_t) \ge -\ln\left(\frac{e^{\eta}+e^{-\eta}}{2}\right)$$

What value for η? Set it to minimize bound.

$$\text{bound}(\eta) = \frac{\ln p}{\Lambda(\eta)} \qquad \frac{d\,\text{bound}(\eta)}{d\eta} = 0 \Rightarrow \eta = \frac{1}{2}\ln\left(\frac{1+\delta}{1-\delta}\right).$$

Corollary: For Winnow, if $\eta = \frac{1}{2}\ln\left(\frac{1+\delta}{1-\delta}\right)$ then

$$\#\text{ mistakes} = T \le \frac{\ln p}{\eta\delta + \ln\left(\frac{2}{e^{\eta}+e^{-\eta}}\right)} = \frac{2\ln p}{\delta^2}$$

$$x = \left[\begin{array}{cccc|c} 0.8 & 0.1 & \vdots & \vdots & -0.8 \\ 0.6 & \vdots & \vdots & & -0.6 \\ 0.4 & \vdots & \vdots & \vdots & -0.4 \\ -0.2 & \vdots & & & 0.2 \end{array}\right]$$

How to compare to perceptron?

perceptron

$$T \le \frac{1}{\delta^2}$$

additive $w^{(t+1)} = w^{(t)} + \_\_\_$

$\|x_i\|_2 \le 1 \quad \forall i$

$\|w^*\|_2 \le 1$

SVM-like $\left\{\begin{array}{l}\text{additive updates}\\ \ell_2\text{-margins}\end{array}\right.$

Winnow

$$T \le \frac{2\ln p}{\delta^2}$$

very similar

$w^{(t+1)} = w^{(t)} \cdot \_\_\_$

$\|x_i\|_\infty \le 1 \quad \forall i$

$\|w^*\|_1 = 1$

AdaBoost $\left\{\begin{array}{l}\text{multiplicative updates}\\ \ell_1\text{-margins}\end{array}\right.$