CMPUT 654 Fa 23: Theoretical Foundations of Machine Learning **Fall 2023**

Lecture 3: September 12

Lecturer: Csaba Szepesvári *Scribes: Shivam Garg*

3.1 Variance bound for binary classification with constant noise and threshold functions

Let $\mathcal{X} = [0,1]$, $\mathcal{Y} = \{0,1\}$, $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ and for $w \in [0,1]$ let $\ell_w : \mathcal{Z} \to \{0,1\}$ be defined by

$$\ell_w(x,y) = \mathbb{I}\{\mathbb{I}\{x \ge w\} \ne y\}\,.$$

Let

$$o_w(x,y) = \mathbb{I}\{x \notin [w-\varepsilon, w]\}\,.$$

Let P be a distribution over $\mathcal{Z}$. As usual, Pf stands for $\int f dP$. For some $w^* \in [0,1]$, we let P_{w^*} be a distribution over $\mathcal{Z}$ such that if $(X,Y) \sim P_{w^*}$ then X is uniformly distributed over $[0,1]$ and $\mathbb{P}(Y=1|X=x) = 1-p$ if $x \ge w^*$ and $\mathbb{P}(Y=1|X=x) = p$ otherwise.

We consider functions of the form $g_{w,w^*} = \ell_w o_w - \ell_{w^*}$. Note that from this choice, it follows that

$$P_{w^*} g_{w,w^*} \ge -p\varepsilon\,.$$

Indeed, writing $P = P_{w^*}$ to minimize clutter, on the one hand we have

$$P\ell_{w^*} = p\,,$$

and also

$$\min_{w\in[0,1]} P\ell_w o_w \ge \min_{w'\in[0,1]} \min_{w\in[0,1]} P\ell_w o_{w'} = \min_{w'\in[0,1]} P\ell_{w^*} o_{w'} = P\ell_{w^*} o_\varepsilon = p - p\varepsilon = p(1-\varepsilon)\,.$$

Hence,

$$Pg_{w,w^*} = P\ell_w o - P\ell_{w^*} \ge p(1-\varepsilon) - p = -p\varepsilon\,.$$

Our next goal is to find the "smallest possible" $c_0, c_1 \ge 0$ such that for any w, w^*, ε, p,

$$P_{w^*} g^2_{w,w^*} \le c_0^2 + c_1 P_{w^*} g_{w,w^*} \tag{3.1}$$

holds. In fact, we want to minimize first c_0 and keep c_1 bounded.

To minimize clutter, we let $P = P_{w^*}$, $o = o_w$, but these dependencies should still be kept in mind. First, algebra gives

$$\begin{aligned} Pg^2_{w,w^*} &= P\ell_w o - 2P\ell_w \ell_{w^*} o + P\ell_{w^*} \\ &= P\ell_w o - 2P\ell_w \ell_{w^*} o + p\,. \end{aligned}$$

Furthermore,

$$Pg = P\ell_w o - P\ell_{w^*} = P\ell_w o - p\,.$$

Hence, Eq. (3.1) is equivalent to

$$P\ell_w o - 2P\ell_w\ell_{w^*}o + p \le c_0^2 + c_1(P\ell_w o - p)\,,$$

which is equivalent to

$$c_1 p + p \le c_0^2 + (c_1 - 1)P\ell_w o + 2P\ell_w\ell_{w^*}o\,. \tag{3.2}$$

Now, we consider three cases based on the location of w^* relative to $w - \varepsilon$ and w and we will in each case lower bound the value of

$$A = (c_1 - 1)P\ell_w o + 2P\ell_w\ell_{w^*}o\,,$$

while enforcing that

$$c_1 \ge 1\,. \tag{3.3}$$

Case 1: $w^* \le w - \varepsilon$. In this case we have

$$\begin{aligned} P\ell_w o &= (w - w^*)(1 - 2p) + p - \varepsilon(1 - p)\,, \\ P\ell_w\ell_{w^*}o &= w^*(1 - p) - wp + p\,. \end{aligned}$$

Then,

$$A = (c_1 - 1)\left\{(w - w^*)(1 - 2p) + p - \varepsilon(1 - p)\right\} + 2\left\{w^*(1 - p) - wp + p\right\}\,.$$

The coefficient of w^* in the above expression is

$$\alpha = (c_1 - 1)(2p - 1) + 2(1 - p)\,.$$

If $\alpha \le 0$, the value of A is minimized in w^* by choosing $w^* = w - \varepsilon$ and the minimum value is equal to

$$\begin{aligned} A' &= (c_1 - 1)\left\{\varepsilon(1 - 2p) + p - \varepsilon(1 - p)\right\} + 2\left\{(w - \varepsilon)(1 - p) - wp + p\right\} \\ &= (c_1 - 1)p(1 - \varepsilon) + 2\left\{(w(1 - 2p) - \varepsilon(1 - p) + p\right\} \\ &\ge (c_1 - 1)p(1 - \varepsilon) + 2\left\{p - \varepsilon(1 - p)\right\} =: A''\,, \end{aligned}$$

where the last inequality follows since $2w(1 - 2p) \ge 0$. Now, Eq. (3.2) will follow if

$$c_1 p(1 - \varepsilon) + p + c_1 p\varepsilon = c_1 p + p \le c_0^2 + A''.$$

Plugging in the definition of A'', we get that the latter inequality is equivalent to

$$\begin{aligned} \cancel{c_1 p(1 - \varepsilon)} + p + c_1\varepsilon &\le c_0^2 + (\cancel{c_1} - 1)p(1 - \varepsilon) + 2\left\{p - \varepsilon(1 - p)\right\} \\ &= c_0^2 - \cancel{p} + p\varepsilon + \cancel{2}p - 2\varepsilon(1 - p) \\ &= c_0^2 + 3\varepsilon p + p - 2\varepsilon \end{aligned}$$

Reordering gives

$$c_0^2 \ge \varepsilon(c_1 + 2 - 3p)\,. \tag{3.4}$$

Since the $p < 1/2$ by assumption and, as we shall see $c_1 \geq 0$, this constraint can be satisfied.

Finally, simple algebra shows that $\alpha \leq 0$ is equivalent to

$$c_1 \geq 1 + \frac{1-p}{\frac{1}{2} - p}\,, \tag{3.5}$$

where we used that $p < 1/2$.

Case 2: $w - \varepsilon < w^* \leq w$. In this case we have

$$\begin{aligned} P\ell_w o &= p(1-\varepsilon)\,, \\ P\ell_w \ell_{w^*} o &= p(1-\varepsilon)\,. \end{aligned}$$

Then,

$$A = (c_1 - 1)p(1-\varepsilon) + 2p(1-\varepsilon) = c_1 p(1-\varepsilon) + p(1-\varepsilon)\,. \tag{3.6}$$

Now, Eq. (3.2) will follow if $c_1(1-\varepsilon)p + p(1-\varepsilon) + (c_1+1)p\varepsilon \leq c_0^2 + A$. Plugging in the definition of A, we get that the latter inequality is equivalent to

$$\cancel{c_1(1-\varepsilon)p} + \cancel{(1-\varepsilon)p} + (c_1+1)p\varepsilon \leq c_0^2 + \cancel{c_1 p(1-\varepsilon)} + \cancel{p(1-\varepsilon)}\,,$$

which gives

$$c_0^2 \geq (c_1 + 1)\varepsilon p\,. \tag{3.7}$$

Case 3: $w < w^*$. In this case we have

$$\begin{aligned} P\ell_w o &= (w - w^*)(2p-1) + p(1-\varepsilon)\,, \\ P\ell_w \ell_{w^*} o &= (w - w^*)p + p(1-\varepsilon)\,. \end{aligned}$$

Then,

$$A = (c_1 - 1)\left\{(w-w^*)(2p-1) + p(1-\varepsilon)\right\} + 2\left\{(w-w^*)p + p(1-\varepsilon)\right\}\,.$$

Note that here $w - w^* \leq 0$. The coefficient of $w - w^*$ in this expression is

$$\alpha = (c_1 - 1)(2p - 1) + 2p.$$

When $\alpha \leq 0$, the value of A is minimized by choosing the largest possible value for $w - w^*$, which is zero. In this case the value of A is

$$A' = (c_1 - 1)p(1-\varepsilon) + 2p(1-\varepsilon) = c_1 p(1-\varepsilon) + p(1-\varepsilon)\,,$$

which is the value we got in the previous case (cf. Eq. (3.6)). As such, as long as Eq. (3.7) holds, Eq. (3.1) will be satisfied.

To guarantee that $\alpha \leq 0$, we need to have

$$(c_1 - 1)(2p - 1) + 2p \leq 0\,.$$

Thanks to $2p - 1 < 0$, this holds if and only if

$$c_1 \geq 1 + \frac{p}{\frac{1}{2} - p}\,. \tag{3.8}$$

Thus, we derived the following:

Proposition 3.1. *Let $p < 1/2$ and let*

$$c_1 = 1 + \frac{1-p}{\frac{1}{2}-p}\,, \qquad c_0^2 = \varepsilon(c_1+2)\,.$$

Then, for any w, w^* it holds that

$$P_{w^*} g_{w,w^*}^2 \leq c_0^2 + c_1 P_{w^*} g_{w,w^*}\,.$$

Proof. Choosing c_1 as above satisfies both Eqs. (3.5) and (3.8). Furthermore, because $\varepsilon(c_1+2) \geq \varepsilon(c_1+2-3p)$ and $\varepsilon(c_1+2) \geq \varepsilon p(c_1+1)$, choosing $c_0^2 = \varepsilon(c_1+2)$ will satisfy both Eqs. (3.4) and (3.7). $\square$