Learning with Statistical Queries

Sunday, November 7, 2021 7:43 PM

Yuxuan

PAC model ✓

What about errors in the labels?

One model for this:

Random Classification Noise Model

each label is flipped with prob. η.

$$(x, \ell(x)) \quad \ell(x) = \begin{cases} f(x) & \text{w.p. } 1-\eta \\ 1-f(x) & \text{w.p. } \eta \end{cases}$$

Q. Can we still learn the underlying concept?

E.g. suppose we are learning an <u>OR</u>.

of Boolean variables.

Let $p_i = \Pr(f(x)=0 \text{ and } x_i = 1)$

We can let output hypothesis h be

all variables for which $p_i = 0$

(any variable with $p_i = 0$ is in the true OR) (any variable with $p_i = 0$ is ...

and no variables for which $p_i > \frac{\varepsilon}{n}$.

So can we estimate each $p_i \pm \frac{\varepsilon}{2n}$ (say)?

$$p_i = \underbrace{\Pr\left(f(x)=0 \,/\, x_i = 1\right)}_{q_i} \cdot \Pr(x_i = 1)$$

independent of label noise

$$\Pr_{\text{noise } \eta}\left(\ell(x)=0 \,/\, x_i = 1\right) = (1-\eta)\, q_i + \eta\,(1-q_i)$$
$$= \eta + q_i(1-2\eta)$$

So if we have LHS, we can subtract η, divide by $(1-2\eta)$.

suffices to approximate to within $\pm \frac{\varepsilon}{2n}(1-2\eta)$.
which we can do from samples!

Statistical Query Model.

Can ask for expectations of bounded functions of $(x, \ell(x))$ up to additive

can ... up to additive
functions of $(x, \ell(x))$ up to additive
error.

E.g. $\mathbb{E}(x_i / \ell(x) = 1)$

$\mathbb{E}(x_i (1-x_j))$

$$\boxed{\begin{array}{l} \chi : \underset{\text{example}}{X} \times \underset{\text{label}}{\{0,1\}} \longrightarrow [0,1] \quad , \tau > 0 \\ \text{SQ oracle responds with } \mathbb{E}_D(\chi) \pm \tau . \end{array}}$$

χ: poly time computable $\quad \tau \geq \frac{1}{\text{poly}}$

Many (almost all) known learning algorithms can be implemented with SQ.

e.g. gradient descent. $\quad L(w) = \mathbb{E}_{x,y}(\ell(x,y,w))$

$$\nabla_w L(w) = \nabla_w \mathbb{E}_{x,y}(\ell(x,y,w))$$

$$= \mathbb{E}_{x,y}(\nabla_w(\ell(x,y,w))$$

$$= \mathbb{E}_{x,y}\left(\nabla_w (\ell(x,y,w)) \right)$$

SQ.

Thm. SQ-learnable $\Rightarrow$ PAC learnable with Random classification noise.

Estimate $\Pr(\chi(x, f(x)) = 1)$

example
$\chi(x, \ell) = 1$
if $x_i = 1$ and $\ell = 0$.

Let

$\underline{\text{CLEAN}} = \{ x : \chi(x, 0) = \chi(x, 1) \}$

$\underline{\text{NOISY}} = \{ x : \chi(x, 0) \neq \chi(x, 1) \}$

$$\Pr(\chi(x, f(x)) = 1) = \Pr(\chi(x, f(x)) = 1 \text{ and } x \in \text{CLEAN}) + \Pr(\chi(x, f(x)) = 1 \text{ and } x \in \text{NOISY}$$

Can. estimate $\Pr(x \in \text{CLEAN})$ — not affected by noise.
and hence first term.

Also $\Pr(x \in \text{NOISY})$

Also $\Pr(x \in \text{NOISY})$

and $\Pr_{\eta}(\chi(x, f(x)) = 1 \mid x \in \text{NOISY}) \} q$

$= (1-\eta)p + \eta(1-p) = \eta + p(1-2\eta)$

where $p = \Pr(\chi(x, f(x)) = 1 \mid x \in \text{NOISY})$

Hence $p = \dfrac{q - \eta}{1 - 2\eta}$.

Need to estimate p to within $\pm \tau$

So need to estimate q to within $\pm \tau(1-2\eta)$