

# Fourier Learning

Monday, November 1, 2021 6:38 AM

Yannik

Motivated by learning DNF.

Still an open problem to PAC learn DNF or decision trees (lists we can learn).

So we consider special distributions

- uniform
- product
- Gaussian etc

and allow for membership queries,  $\text{N}\cdot\ell$ ,  
"What is the label of  $x$ ?"

Assume  $x \in \{-1, 1\}^n$   $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$

Note that any such  $f \in \{-1, 1\}^{2^n}$  (as a table)

Defines inner product of  $f$  and  $g$  with respect to distribution  $D$  as

$$\langle f, g \rangle = \sum D(x) f(x) g(x)$$

$$\langle f, g \rangle_D = \sum_x D(x) f(x) g(x)$$

$$\langle f, f \rangle_D = \|f\|_D^2 = 1 \quad (\text{since } f(x)^2 = 1)$$

Viewing  $f$  as a vector, the standard basis is  $e_1, e_2, \dots, e_{2^n}$ .

But we can use any basis and write  
 $f(x) = \sum_v \langle f, v \rangle v$  where  $\{v\}$  is an orthonormal basis.

What's an interesting basis?

The set of parity functions?

$\forall S \subseteq [n], X_S(x) = \prod_{i \in S} x_i$   $2^n$  functions.

$$\langle X_S, X_S \rangle_D = 1$$

$$\langle X_S, X_T \rangle_D = \mathbb{E}_D \left( \prod_{i \in S} X_i \prod_{j \in T} X_j \right) = 0$$

for a product distribution  $D$ .

for a product distribution  $\hookrightarrow$   
Hence  $\{x_s\}$  is an orthonormal basis!

So any f can be written as

$$f(x) = \sum_s \hat{f}_s \chi_s(x) \quad \text{where } \hat{f}_s = \langle f, \chi_s \rangle$$

$$\text{Thm 1. (Parseval)} \quad \langle f, f \rangle_D = \langle \hat{f}, \hat{f} \rangle$$

Thm 2 (Plancherel)  $\langle f, g \rangle_D = \langle \hat{f}, \hat{g} \rangle.$

$$\begin{aligned}
 \underline{\text{PF}} \cdot \sum_x \mathcal{D}(x) \sum_s \hat{f}_s \chi_s(x) &\sum_T \hat{g}_T \chi_T(x) \\
 = \sum_{S,T} \hat{f}_S \hat{g}_T \mathbb{E}_{\mathcal{D}} (\chi_S(x) \chi_T(x)) \\
 = \sum_S \hat{f}_S \hat{g}_S &= \langle \hat{f}, \hat{g} \rangle.
 \end{aligned}$$

A decision tree is a Boolean function  $f$ .  
We want to learn  $f$  by approximating all

We want to learn + <sup>log n</sup> approx of its significant Fourier coefficients  $\hat{f}_s$ .

Our approx is  $\hat{g}$ .

$$\Pr_D(g(x) \neq f(x)) \leq \mathbb{E}_D((f(x) - g(x))^2) = \sum_s (\hat{f}_s - \hat{g}_s)^2.$$

we will learn all  $\hat{f}_s$  for which  $\hat{f}_s \geq \tau$ .

Note.  $\sum_s \hat{f}_s^2 = 1 \Rightarrow |\hat{f}_s| \leq 1$

Lemma. If a decision tree has  $m$  leaves

then  $\|f\|_1 = \sum |\hat{f}_s| \leq 2m+1$ .

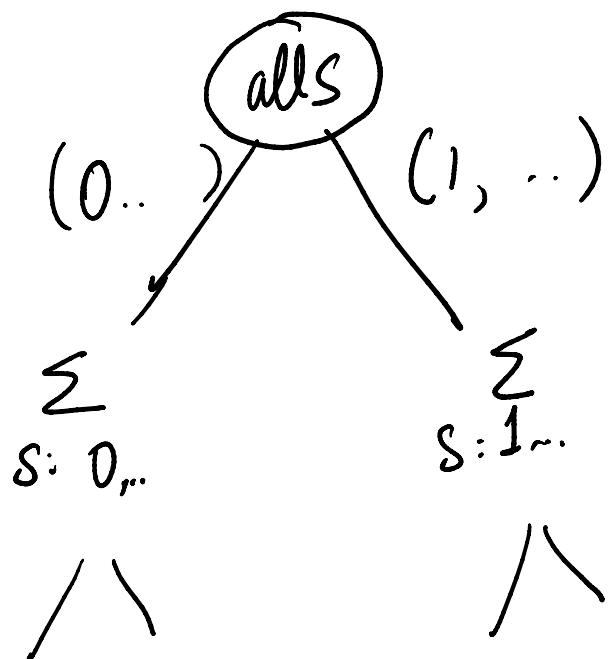
Thm. If we learn all  $\hat{f}_s \geq \frac{\varepsilon}{\|f\|_1}$

then  $\|\hat{f}_s - \hat{g}_s\|^2 \leq \varepsilon$ .

Pf.  $\|\hat{f}_s - \hat{g}_s\|^2 = \sum_{s: |\hat{f}_s| \leq \frac{\varepsilon}{\|f\|_1}} \hat{f}_s^2 \leq \sum_s |\hat{f}_s| \cdot \frac{\varepsilon}{\|f\|_1}$

$$\Rightarrow \|f_S\| \leq \frac{\epsilon}{\|f_1\|} \rightarrow = \epsilon.$$

How to learn all large Fourier Coefficients.



At each node  
estimate whether

$$\sum \hat{f}_s^2 \geq T.$$

S: prefix  $\alpha$

$$\text{width} \leq \frac{1}{T}, \text{depth} \leq n$$

$$\# \text{nodes} \leq \frac{n}{T}.$$

How to estimate

$$\sum_{S \in S_\alpha} \hat{f}_s^2 ?$$

Suppose  $\alpha = (0, \underbrace{\dots}_k 0)$ .

$$r_{\min} - \hat{\gamma}^2 \neq (f(yx) f(zx)).$$

$$\text{Claim: } \sum_{S_\alpha} \hat{f}_S^2 = \mathbb{E}_{\substack{x \sim \{0,1\}^{n-k}, y, z \sim \{0,1\}^k}} (f(yx) f(zx))$$

Suppose  $f$  is a parity function

if  $f$  agrees with  $\alpha$ , then  $f(yx) = f(zx)$  so we get 1.

else  $\Pr(f(yx) = f(zx)) = \frac{1}{2} \Rightarrow$  we get 0.

Any  $f$  can be written as a weighted sum of parities. So,

$$f = \sum_U \hat{f}_U \chi_U$$

$$\begin{aligned} \mathbb{E}(f(yx) f(zx)) &= \mathbb{E}\left(\sum_U \hat{f}_U \chi_U(yx) \sum_V \hat{f}_V \chi_V(zx)\right) \\ &= \sum_{U,V} \hat{f}_U \hat{f}_V \underbrace{\mathbb{E}(\chi_U(yx) \chi_V(zx))}_{=0 \text{ if } U \neq V} \\ &= \sum_U \hat{f}_U^2 \underbrace{\mathbb{E}(\chi_U(yx) \chi_U(zx))}_{=0 \text{ if } U \text{ does not }} \end{aligned}$$

$$= \sum_{U \in U_\alpha} \hat{f}_U^2$$

= 0 if  $U$  does not agree with  $\alpha = (0, \cdot)$

What about general  $\alpha$ ?

Lemma:  $\sum_{S \in S_\alpha} \hat{f}_S^2 = \mathbb{E} (f(y) f(z) \chi_\alpha(y) \chi_\alpha(z))$

$x \sim \xi_0, \mathcal{I}^{n-k}$   
 $y, z \sim \xi_0, \mathcal{I}^k$ .

---