Computer Science 294 Lecture 16 Notes

Daniel Raban

March 9, 2023

1 Sheppard's Lemma and Noise Sensitivity of LTFs

1.1 Recap: Noise stability of the majority function

Last time, we saw the noise stability of the majority function.

Theorem 1.1. *For all $\rho \in [-1, 1]$,*

$$\mathrm{Stab}_\rho(\mathrm{MAJ}_n) \xrightarrow{n \to \infty} \frac{2}{\pi} \arcsin \rho.$$

The only missing piece for us was Sheppard's lemma. We will also see that the majority is the stablest among LTFs via the following result concerning noise stability.

Theorem 1.2 (Peres)**.** *For every LTF f, $\mathrm{NS}_\delta(f) = O(\sqrt{\delta})$.*

1.2 Multivariate Gaussians

Definition 1.1. A random variable $Z = (Z_1, \dots, Z_n) \in \mathbb{E}^n$ is called a **multivariate Gaussian (MVG)** if for any linear function $\ell : \mathbb{R}^n \to \mathbb{R}$, $\ell(Z)$ is a univariate Gaussian.

Remark 1.1. We can't just say that the marginal distributions are Gaussians because we could have $Z_1 \sim N(0, 1)$ and

$$Z_2 = \begin{cases} Z_1 & \text{with probability } 1/2 \\ -Z_1 & \text{with probability } 1/2. \end{cases}$$

These are both Gaussian, but $Z_1 + Z_2$ is not Gaussian.

A multivariate Gaussian distrbituion is parameterized by an n -dimensional mean vector $\mu \in \mathbb{R}^n$ and an $n \times n$ covariance matrix $\mathbb{R}^{n \times n}$. We write

$$Z \sim N(\mu, \Sigma), \qquad \mathbb{E}[Z_i] = \mu_i, \qquad \mathrm{Cov}(Z_i, Z_j) = \mathbb{E}[(Z_i - \mu_i)(Z_j - \mu_j)].$$

The **standard multivariate Gaussian** is $Z \sim N(0, I_n)$. Alternatively, $Z_1, \dots, Z_n$ are iid stndard Gaussians.

Proposition 1.1. *For any fixed $\mu \in \mathbb{R}^n$, X sim$\mathbb{N}(\mu, \Sigma)$ if and only if there exists a linear transformation $A \in \mathbb{R}^{n \times k}$ such that $x = AZ + \mu$, where $Z \sim N(0_k, I_k)$.*

Proof. We only prove one direction. Suppose that $X = AZ + \mu$, where $Z \sim N(0_k, I_k)$. Then $\ell(X) = \ell'(Z)$ is a linear combination of the marginals of Z, so $\ell(Z)$ is a univariate Guassian random variable. More explicitly,

$$\mathbb{E}[X_i] = \mathbb{E}[(AZ)_i + \mu_i] = \mathbb{E}\left[\sum_j A_{i,j} Z_j + \mu_i\right] = \mu_i,$$

$$\mathrm{Cov}(X_i, X_j) = (AA^\top)_{i,j} \implies \Sigma = AA^\top.$$

□

One of the most important properties of the standard multivariate Gaussian distribution is that it is rotationally symmetric. Let $Z = (Z_1, \dots, Z_n) \sim N(0_n, I_n)$, which has joint probability density function

$$f_{Z_1,\dots,Z_n}(z_1, \dots, z_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2} = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-(z_1^2 + \cdots + z_n^2)/2}$$

Note that this probability density function only depends on $z_1^2 + \cdots + z_n^2$; that is, it only depends on the radius of the vector $(z_1, \dots, z_n)$.

Corollary 1.1. *Let $Z = (Z_1, \dots, Z_n) \sim N(0_n, I_n)$. Then $\frac{Z}{\|Z\|_2}$ is a uniformly random unit vector in $\mathbb{R}^n$.*

1.3 Sheppard's lemma

Definition 1.2. Two Gaussians (Z_1, Z_2) are **ρ-correlated** if

$$(Z_1, Z_2) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right).$$

Equivalently,

$$Z = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix},$$

where $X \sim N(0_2, I_2)$.

Lemma 1.1 (Sheppard)**.** *Let (Z_1, Z_2) be ρ-correlated Gaussians. Then*

$$\mathbb{P}(\mathrm{sgn}(Z_1) \neq \mathrm{sgn}(Z_2)) = \frac{1}{\pi} \arccos \rho.$$

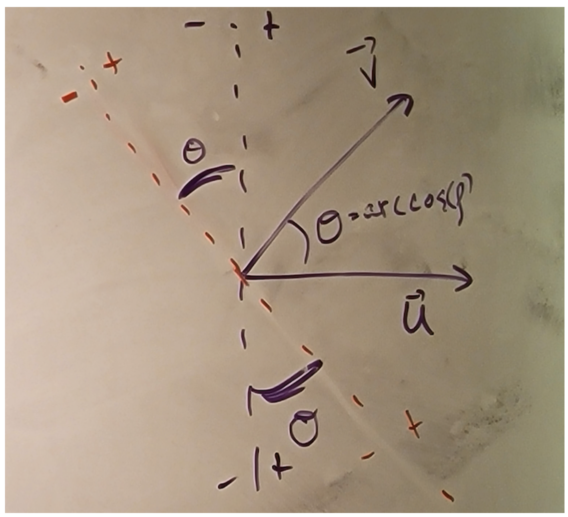
Proof. Let X_1, X_2 be independent standard Gaussians, and let

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}.$$

Note that if the first row of this matrix is u and the second row of this matrix is v, then $Z_1 = \langle u, X \rangle$ and $Z_2 + \langle v, X \rangle$. Then

$$\begin{aligned} \mathbb{P}(\operatorname{sgn}(Z_1) \neq \operatorname{sgn}(Z_2)) &= \mathbb{P}(\operatorname{sgn}(\langle u, X \rangle) \neq \operatorname{sgn}(\langle v, X \rangle)) \\ &= \mathbb{P}(\operatorname{sgn}(\langle u, \tfrac{X}{\|X\|} \rangle) \neq \operatorname{sgn}(\langle v, \tfrac{X}{\|X\|} \rangle)), \end{aligned}$$

where $\frac{X}{\|X\|}$ is a uniformly distributed unit vector in $\mathbb{R}^2$. Noe $\rho = \langle u, v \rangle = \cos\theta$, so $\theta = \arccos\rho$, where θ is the angle between the unit vectors u and v. Now we can check which unit vectors in $\mathbb{R}^2$ give a different sign for inner product with u vs inner product with v:



Vectors in arcs of total angle 2θ will make these signs disagree, so we get

$$\mathbb{P}(\operatorname{sgn}(Z_1) \neq \operatorname{sgn}(Z_2)) = \frac{2\theta}{2\pi} = \frac{\arccos\rho}{\pi}.$$ □

1.4 Noise sensitivity of linear threshold functions

Recall that if $f : \{\pm 1\}^n \to \{\pm 1\}$ and $\delta \in [0, 1]$, the **noise sensitivity** of f (at noise rate δ) is given as follows.

- Pick $X \sim \{\pm 1\}^n$ uniformly at random.
- Pick Y by flipping each bit in X independently with probability δ.

$$\mathrm{NS}_\delta(f) = \mathbb{P}(f(X) \neq f(Y)).$$

We saw before that the noise stability is related to the noise sensitivity by

$$\mathrm{NS}_\delta(f) = \frac{1}{2} - \frac{1}{2}\,\mathrm{Stab}_{1-2\delta}(f).$$

We are interested in proving the following theorem.

Theorem 1.3 (Peres)**.** *If f is an LTF, then for all δ, $\mathrm{NS}_\delta \leq O(\sqrt{\delta})$.*

First, we prove the following claim.

Lemma 1.2. *If $f : \{\pm 1\}^n \to \{\pm 1\}$ is an LTF, then $\mathbb{I}(f) \leq \mathbb{I}(\mathrm{MAJ}_n) \leq \sqrt{n}$.*

Proof of lemma. Let $f = \mathrm{sgn}(a_0 + a_1 x_1 + \cdots + a_n x_n)$. If f is monotone, then

$$\mathbb{I}(f) = \mathrm{Eff}(f) \leq \mathrm{Eff}(\mathrm{MAJ}_n) \leq \sqrt{n}.$$

Not every LTF is monotone, but up to flipping the value of the inputs, we can write f as a monotone function g. For example, if we look at the non-monotone function

$$f(x) = \mathrm{sgn}(x_1 - 3x_2 + 5x_3 + 7x_4 - x_5),$$

we can write it as

$$f(x) = \mathrm{sgn}(x_1 + 3(-x_2) + 5x_3 + 7x_4 + (-x_5)) = g(x_1, -x_2, x - 3, x_4, x_5),$$

where

$$g(y) = \mathrm{sgn}(x_1 + 4x_2 + 5x_3 + 7x_4 + x_5).$$

Then

$$\mathbb{I}(f) = \sum_{i=1}^n \mathrm{Inf}_i(f) = \sum_{i=1}^n \mathrm{Inf}_i(g) = \mathrm{Eff}(g) \leq \mathrm{Eff}(\mathrm{MAJ}_n) \leq \sqrt{n}. \qquad \square$$

Remark 1.2. This proof actually applies to any **unate** function f, not just LTFs.

Theorem 1.4. *Let $\mathcal{C}$ be any class of functions that is closed under projections (LTFs are such a class). Suppose that for all n bit functions in this class, $\mathbb{I}(f) \leq A(n)$. Then for every positive integer m,*

$$\mathrm{NS}_{1/m}(f) \leq \frac{1}{m} A(m).$$

What is a projection? First, let's see how this implies Peres' theorem.

Proof of Peres' theorem. Let $\delta > 0$. We want to show that $\mathrm{NS}_\delta \leq O(\sqrt{\delta})$. Take $m = \lceil 1/\delta \rceil$. Then $\mathrm{NS}_\delta(f) \leq \mathrm{NS}_{1/m}(f)$ because noise sensitivity is monotone in the parameter. So taking $A(n) = \sqrt{n}$ in the theorem, we get

$$\mathrm{NS}_\delta(f) \leq \frac{1}{m}\sqrt{m} = \frac{1}{\sqrt{m}} = O(\sqrt{\delta}). \qquad \square$$

What is a projection? A restriction takes $f(y_1, \ldots, y_n)$ and replaces each y_i with -1, $+1$, or y_i.

Definition 1.3. A **projection** takes $f(y_1, \ldots, y_n)$ and replaces y_i with -1, $+1$, z_1,..., z_n, $-z_1$,..., or $-z_n$.

Example 1.1. If $f(y) = \mathrm{sgn}(a_0 + a_1 y_1 + a_2 y_2 + a_3 y_3 + a_4 y_4)$, then we can replace y_1, y_2 by z_2, y_3 by z_1, and y_4 by $-z_1$. Then the projection gives

$$g(z_1, z_2) = \mathrm{sgn}(a_0 + (a_1 + a_2) z_2 + (a_3 - a_4) z_1).$$

Many classes of functions are closed under projections, so this is not a very strong assumption.

Proof of theorem. Let $f \in \mathcal{C}$. Then, letting X be uniform and letting Y be obtained by flipping each bit with probability $1/m$,

$$\mathrm{NS}_{1/m}(f) = \mathbb{P}(f(X) \neq f(Y)).$$

Let's see another way to sample X and Y:

Step 1: Pick X uniformly at random.

Step 2: Partition the n coordinates to m parts uniformly at random, giving a map $\pi : [n] \to [m]$.

Step 2.5: Pick $Z \in \{\pm 1\}^m$ uniformly at random. Attain X' by flipping each i part if $Z_i = -1$.

Step 3: Pick a random part $j \in [m]$ and attain y by flipping all coordinates in the j-th part in X'.

This gives the desired distribution for (X', Y):

$$\begin{aligned} \mathrm{NS}_{1/m}(f) &= \mathbb{P}_{X,\pi,Z,j}(f(X') \neq f(Y)) \\ &= \mathbb{E}_{X,\pi}[\mathbb{P}_{Z,j}(f(X') \neq f(Y))] \end{aligned}$$

Define $g_{x,\pi}(z) = f(x_1 \cdot z_{\pi(1)}, x_2 \cdot z_{\pi(2)}, \ldots, x_n z_{\pi(n)})$. This is a projection of f, so $g_{X,\pi} \in \mathcal{C}$.

$$\begin{aligned} &= \mathbb{E}_{X,\pi}[\mathbb{P}_{Z,j}(g_{X,\pi}(Z) \neq g_{X,\pi}(Z^{\oplus j}))] \\ &= \mathbb{E}_{X,\pi}\left[\frac{1}{m}\sum_{j=1}^{m} \mathrm{Inf}_j(g_{X,\pi})\right] \\ &\leq \frac{A(m)}{m}. \end{aligned}$$

□

This tells us that LTFs are ε-concentrated up to degree $O(1/\varepsilon^2)$. So the LMN lemma tells us that we can learn LTFs in $n^{o(1/\varepsilon^2)}$ time.

Later, we will prove the "majority is stablest" theorem. The Fourier representation for noise stability is

$$\mathrm{Stab}_\rho(f) = \sum_S \widehat{f}(S)^2 \rho^{|S|} = \rho W^1(f) + \rho^2 W^2(f) + \cdots .$$

So for small ρ, we can understand the noise stability by studying

$$W^1(f) = \sum_{i=1}^n \widehat{f}(\{i\})^2.$$

Analyzing this with the Berry-Esseen theorem will give the following.

Theorem 1.5 (2/π theorem)**.** *Let* $f : \{\pm 1\}^n \to \{\pm 1\}$ *satisfy* $|\mathrm{Eff}_i(f)| \leq \varepsilon$. *Then*

$$W^1(f) \leq W^1(\mathrm{MAJ}_n) + O(\varepsilon) \approx \frac{2}{\pi}.$$