CS 6817 HW 2

2.38, 2.46, 2.56 (a) and (b)

2.38

We use this formula of stability:

$$Stab_\rho[f] = \sum_{S\subseteq[n]} \rho^{|S|} \cdot \hat{f}(s)^2$$

The formula for Fourier coefficients of the Tribes function is given in the textbook on page 97:

There are s tribes with w members each. For each $S \subseteq [ws]$, we write it as $(S_1, \dots, S_s)$, where S_i is the intersection of S with the i-th tribe:

$$\widehat{Tribes}_{w,s}(S) = \begin{cases} 2(1-2^{-w})^s - 1, & if\ S = \emptyset \\ 2(-1)^{k+|S|}2^{-kw}(1-2^{-w})^{s-k}, & if\ k = |\{i: S_i \neq \emptyset\}| > 0 \end{cases}$$

Since the we need $\hat{f}(s)^2$ in the stability formula:

$$\widehat{Tribes}_{w,s}(S)^2 = \begin{cases} 4(1-2^{-w})^{2s} - 4(1-2^{-w})^s + 1, & if\ S = \emptyset \\ 2^{-2kw+2}(1-2^{-w})^{2s-2k}, & if\ k = |\{i: T_i \neq \emptyset\}| > 0 \end{cases}$$

If we define $k = |\{i: T_i \neq \emptyset\}|$, then the above could be rewritten as

$$\widehat{Tribes}_{w,s}(S)^2 = \begin{cases} 2^{-2kw+2}(1-2^{-w})^{2s-2k} - 4(1-2^{-w})^s + 1, & if\ S = \emptyset \\ 2^{-2kw+2}(1-2^{-w})^{2s-2k}, & otherwise \end{cases}$$

Since $k = 0$ when $S = \emptyset$. Notice both instances have $2^{-2kw+2}(1-2^{-w})^{2s-2k}$ in common.

So,

$$Stab_\rho[Tribes] = 1 - 4(1-2^{-w})^s + \sum_{S\subseteq[ws]} \rho^{|S|} \cdot 2^{-2kw+2}(1-2^{-w})^{2s-2k}$$

We want to simplify something:

$$\begin{aligned} 2^{-2kw+2}(1-2^{-w})^{2s-2k} &= 4(2^w)^{-2k}(1-2^{-w})^{-2k}(1-2^{-w})^{2s} \\ &= 4\left(2^w(1-2^{-w})\right)^{-2k}(1-2^{-w})^{2s} \\ &= 4(2^w-1)^{-2k}(1-2^{-w})^{2s} \end{aligned}$$

So,

$$\boxed{Stab_\rho[Tribes] = -4(1-2^{-w})^s + \sum_{S\subseteq[ws]} 4\rho^{|S|} \cdot \frac{(1-2^{-w})^{2s}}{(2^w-1)^{2k}}}$$

The Piazza hint says to take asymptotics into account. I'm not sure, but $\lim_{k\to\infty} \frac{(1-2^{-w})^{2s}}{(2^w-1)^{2k}} = 0$ so maybe $\boxed{Stab_\rho[Tribes] = -4(1-2^{-w})^s}$

2.46

The Mean Value Theorem for a function f continuous in the domain interval $[a,b]$ is:

$$\exists c \in (a,b) \text{ such that } f'(c) = \frac{f(b)-f(a)}{b-a}$$

We propose that

$$\begin{gathered} a = \rho - \epsilon \\ b = \rho \\ b - a = \epsilon \\ f(a) = Stab_{\rho-\epsilon}[f] \\ f(b) = Stab_\rho[f] \end{gathered}$$

So, for some $c \in (\rho-\epsilon, \rho)$:

$$\frac{Stab_p[f] - Stab_{p-\epsilon}[f]}{\epsilon} = \frac{\big(f(b)-f(a)\big)}{b-a} = f'(c)$$

Which means we have to show

$$Stab'(c) \le \frac{1}{1-\rho} Var(f)$$

We differentiate $Stab_c(f)$ with respect to the noise parameter c:

$$Stab_c'(f) = \frac{d}{dc} Stab_c(f) = \frac{d}{dc} \sum_{S\subseteq[n]} c^{|S|}\cdot \hat{f}(s)^2 = \sum_{S\subseteq[n]} \frac{d}{dc}\left(c^{|S|}\cdot \hat{f}(s)^2\right)$$
$$= \sum_{S\subseteq[n]} |S| c^{|S|-1}\cdot \hat{f}(s)^2 = \sum_{S\subseteq[n]\ s.t.\ S\neq\emptyset} |S| c^{|S|-1}\cdot \hat{f}(s)^2$$

Because when $S=\emptyset$, the expression inside the summation is 0 anyway.

Now we set $1-d = c$ and employ knowledge from the previous exercise, 2.45:

$$|S| c^{|S|-1} \le \frac{1}{1-c}$$

Which makes

$$Stab_c'(f) \le \frac{1}{1-c} \cdot \sum_{S\subseteq[n]\ s.t.\ S\neq\emptyset} \hat{f}(s)^2$$

Now we need to prove

$$\frac{1}{1-\rho}Var(f) = \frac{1}{1-\rho} \cdot \sum_{S\subseteq[n]\ s.t.\ S\neq\emptyset} \hat{f}(s)^2 \geq Stab'_c(f)$$

Well, since $c < \rho$ and $\rho \in [0,1)$,

$$\frac{1}{1-\rho} > \frac{1}{1-c}$$

Thus,

$$\frac{1}{1-\rho}Var(f) \geq Stab'(c)$$

QED.

2.56

(a)

Ideally, we want $f_1(y^{(1)}) = f_2(y^{(2)})$, which means ideally, $\mathbb{E}\left(f_1(y^{(1)}) \cdot f_2(y^{(2)})\right) = 1$, since $-1^2 = 1^2 = 1$. So, we want to maximize $\mathbb{E}\left(f_1(y^{(1)})f_2(y^{(2)})\right) = < f_1(y^{(1)}), f_2(y^{(2)}) >$.

$$< f_1(y^{(1)}), f_2(y^{(2)}) > \ = \ < \sum_{S\subseteq[n]} \widehat{f_1}(S)\rho^{|S|}X_S\,, \sum_{T\subseteq[n]} \widehat{f_2}(T)\rho^{|T|}X_T >$$
$$= \sum_{S,T\subseteq[n]} (\widehat{f_1}(S)\rho^{|S|})(\widehat{f_2}(T)\rho^{|T|}) < X_S, X_T >$$

Due to the noise operator formula.

Now we use the Cauchy-Schwarz inequality. We set

$$a_i = \widehat{f_1}(S)\rho^{|S|}$$
$$b_i = (\widehat{f_2}(S)\rho^{|S|}) < X_S, X_T >$$

So,

$$\left(\sum_{S\subseteq[n]} (\widehat{f_1}(S)\rho^{|S|})^2\right)\left(\sum_{T\subseteq[n]} (\widehat{f_2}(T)\rho^{|T|} < X_S, X_T >)^2\right)$$
$$\geq \left(\sum_{S,T\subseteq[n]} (\widehat{f_1}(S)\rho^{|S|})(\widehat{f_2}(T)\rho^{|T|}) < X_S, X_T >\right)^2$$

Left side:

$$\left(\sum_{S\subseteq[n]}\left(\widehat{f_1}(S)\rho^{|S|}\right)^2\right)\left(\sum_{T\subseteq[n]}\left(\widehat{f_2}(T)\rho^{|T|}<\mathrm{X}_S,\mathrm{X}_T>\right)^2\right)=\left(\sum_{S\subseteq[n]}\widehat{f_1}(S)^2\rho^{2|S|}\right)\left(\sum_{T\subseteq[n]}\widehat{f_2}(T)^2\rho^{2|T|}\right)$$
$$=\mathbb{E}\left(\rho^{2|S|}\right)\mathbb{E}\left(\rho^{2|T|}\right)$$

Because

$$<\mathrm{X}_S,\mathrm{X}_T>^2=1$$

And because of Parseval's identity.

Now, we want to maximize

$$\mathbb{E}\left(\rho^{2|S|}\right)\mathbb{E}\left(\rho^{2|T|}\right)$$

Because $\rho\in(0,1)$, $\rho^{2|S|},\rho^{2|T|}\in(0,1)$. $\rho^{2|S|}$ and $\rho^{2|T|}$ are maximized when $|S|$ and $|T|$ are as small as possible. And since we can't have a trivial solution, they have to be greater than 0, so $|S|=|T|=1$, **so $\boldsymbol{f_1}$ and $\boldsymbol{f_2}$ must be dictator functions**:

$$f_1\left(y^{(1)}\right)=y_i^{(1)} \text{ or } f_1\left(y^{(1)}\right)=-y_i^{(1)} \text{ for some } y_i^{(1)}$$

and

$$f_2\left(y^{(2)}\right)=y_j^{(2)} \text{ or } f_2\left(y^{(2)}\right)=-y_j^{(2)} \text{ for some } y_j^{(2)}$$

And we need

$$f_1\left(y^{(1)}\right)\cdot f_2\left(y^{(2)}\right)$$

To be as large as possible.

We want

$$\mathbb{E}\left(f_1\left(y^{(1)}\right)\cdot f_2\left(y^{(2)}\right)\right)$$

to be greater than 0 because we squared it during the Cauchy-Schwarz process. To make this as likely as possible, we need the two dictator functions to be dictated by the same bit i because we have no guarantees about choosing two different bits i and j since they're independent of each other; and we need them to be either both positive dictators or both negative dictators because $1\cdot 1=-1\cdot -1=1$, whereas $1\cdot -1=-1$, and because $f_1\left(y^{(1)}\right)$ and $f_2\left(y^{(2)}\right)$ are actually more likely to be equal than unequal:

$$\mathbb{P}\left(f_1(y^{(1)}) = f_2(y^{(2)})\right) = \left(\frac{1}{2}+\frac{1}{2}\rho\right)^2 + \left(\frac{1}{2}-\frac{1}{2}\rho\right)^2 = \frac{1}{2}+\frac{1}{2}\rho^2$$

(Because they're equal when they both get x_i right or both get x_i wrong)

$$\mathbb{P}\left(f_1(y^{(1)}) \neq f_2(y^{(2)})\right) = 2\left(\frac{1}{2}-\frac{1}{2}\rho\right)^2 = \frac{1}{2}-\frac{1}{2}\rho^2$$

(Because they're unequal when one of them gets x_i right and the other gets it wrong)

Thus, the functions themselves need to be:

$$f_1(y^{(1)}) = y_i^{(1)} \text{ and } f_2(y^{(2)}) = y_i^{(2)} \text{ for some bit } i$$

or

$$f_1(y^{(1)}) = -y_i^{(1)} \text{ and } f_2(y^{(2)}) = -y_i^{(2)} \text{ for some bit } i$$

i.e. dictated by the same bit, and both multiplied by $+1$ or -1.

(b)

We strive for a similar objective as in part a of this problem, but this time it's 3 objectives:

$$\text{Maximize } \left(f_1(y^{(1)})f_2(y^{(2)})\right)$$
$$\text{Maximize } \left(f_1(y^{(1)})f_3(y^{(3)})\right)$$
$$\text{Maximize } \left(f_2(y^{(2)})f_3(y^{(3)})\right)$$

We do the steps we did above for each of these, giving two possible results for each:

$$f_1(y^{(1)}) = y_i^{(1)} \text{ and } f_2(y^{(2)}) = y_i^{(2)} \text{ for some bit } i$$

or

$$f_1(y^{(1)}) = -y_i^{(1)} \text{ and } f_2(y^{(2)}) = -y_i^{(2)} \text{ for some bit } i$$

and

$$f_1(y^{(1)}) = y_j^{(1)} \text{ and } f_3(y^{(3)}) = y_j^{(3)} \text{ for some bit } j$$

or

$$f_1(y^{(1)}) = -y_j^{(1)} \text{ and } f_3(y^{(3)}) = -y_j^{(3)} \text{ for some bit } j$$

and

$$f_2(y^{(2)}) = y_k^{(2)} \text{ and } f_3(y^{(3)}) = y_k^{(3)} \text{ for some bit } k$$

or

$$f_2(y^{(2)}) = -y_k^{(2)} \text{ and } f_3(y^{(3)}) = -y_k^{(3)} \text{ for some bit } k$$

And because we have to link the three together, i has to equal j has to equal k. QED.