





$$(1-\epsilon)^m \leq e^{-\epsilon m}$$

לנקיטה

$$\Rightarrow e^{-\epsilon m} \leq \delta \Rightarrow \ln(e^{-\epsilon m}) \leq \ln(\delta)$$

$$\Rightarrow -\epsilon m \leq \ln(\delta) \Rightarrow m \geq \frac{\ln(\frac{1}{\delta})}{\epsilon}$$

$$m \geq \frac{\ln(\frac{1}{\delta})}{\epsilon} \quad \text{לנקיטה } m \text{ מספיק גדול}$$

$$P_{S \sim D} [L_D(h_S) \leq \epsilon] \geq 1 - \delta, \quad m_{\text{req}}(\epsilon, \delta) \leq \frac{\log(\frac{1}{\delta})}{\epsilon}$$

3. נניח  $V_{\text{dim}}(H) = d$  ו-  $C \subseteq X$  קבוצה קטנה.  $d$  הנקרא  $d$  ממדינות של  $H$ .

כל  $h \in H$  יהיה  $h$  מתואר על ידי  $d$  פרמטרים. כלומר,  $h$  יכול להיות  $h = (x_1, \dots, x_d)$  עבור  $x_i \in \mathbb{R}$ .

$$d \leq \log_2 |H| \Leftrightarrow 2^d \leq |H|$$

$$\Rightarrow V_{\text{dim}}(H) \leq \log_2 |H| \quad \square$$

4. נניח  $V_{\text{dim}}(H_{\text{arity}}) \leq n$  ו-  $C \subseteq X$  קבוצה קטנה.  $n$  הנקרא  $n$  ממדינות של  $H_{\text{arity}}$ .

נניח  $C = \{x_1, \dots, x_n\}$  ו-  $x_i \in \mathbb{R}^n$ . נניח  $C$  קבוצה קטנה.

נניח  $C$  קבוצה קטנה. נניח  $C$  קבוצה קטנה.

נניח  $C$  קבוצה קטנה. נניח  $C$  קבוצה קטנה.

נניח  $C = \{e_1, \dots, e_n\}$  ו-  $e_i \in \mathbb{R}^n$ . נניח  $C$  קבוצה קטנה.

נניח  $C$  קבוצה קטנה. נניח  $C$  קבוצה קטנה.

נניח  $C$  קבוצה קטנה. נניח  $C$  קבוצה קטנה.

נניח  $C$  קבוצה קטנה. נניח  $C$  קבוצה קטנה.

$$I = \emptyset \Leftrightarrow \gamma = [0, \dots, 0]^n$$

$$I = \{1\} \Leftrightarrow \gamma = [1, 0, \dots, 0]^n$$

$$I = \{1, \dots, n\} \Leftrightarrow \gamma = [0, \dots, 0, 1]^n$$



$\mathcal{H}$  הוא קבוצת נקודות  $x_1, x_2, \dots, x_n$  הנמצאת ב- $\mathbb{R}^n$  (המרחב  $n$ -ממדי).  
 $V_{dim}(\mathcal{H}) = n$  אם ורק אם  $(e_1, \dots, e_n)$  היא בסיס.  
 $V_{dim}(\mathcal{H}_{k-interval}) = 2k - 1$  (60)

$V_{dim}(\mathcal{H}_{n-interval}) \leq 2n$  : נניח  $x_1 < x_2 < \dots < x_{2n+1}$  אזי  
 נקודה  $x_i$  היא נקודה ב- $\mathbb{R}^n$  הממוקמת על קו.

נניח  $x_1 < x_2 < \dots < x_{2n+1}$  אזי  $x_i$  היא נקודה ב- $\mathbb{R}^n$  הממוקמת על קו.  
 נניח  $x_1 < x_2 < \dots < x_{2n+1}$  אזי  $x_i$  היא נקודה ב- $\mathbb{R}^n$  הממוקמת על קו.  
 נניח  $x_1 < x_2 < \dots < x_{2n+1}$  אזי  $x_i$  היא נקודה ב- $\mathbb{R}^n$  הממוקמת על קו.

נניח  $x_1 < x_2 < \dots < x_{2n+1}$  אזי  $x_i$  היא נקודה ב- $\mathbb{R}^n$  הממוקמת על קו.

$V_{dim}(\mathcal{H}_{k-interval}) \leq 2k$  : נניח  $x_1 < x_2 < \dots < x_{2k+1}$  אזי

נניח  $x_1 < x_2 < \dots < x_{2k+1}$  אזי  $x_i$  היא נקודה ב- $\mathbb{R}^n$  הממוקמת על קו.

נניח  $x_1 < x_2 < \dots < x_{2k+1}$  אזי  $x_i$  היא נקודה ב- $\mathbb{R}^n$  הממוקמת על קו.

$V_{dim}(\mathcal{H}_{n-interval}) = 2n$

נניח  $x_1 < x_2 < \dots < x_{2n+1}$  אזי  $x_i$  היא נקודה ב- $\mathbb{R}^n$  הממוקמת על קו.







[illegible][illegible]

7. (נדר) אגנוסטיק פאקט - אגנוסטיק פאקט

$$m \geq m_4(\varepsilon, \delta) \quad \text{כאשר} \quad \delta, \varepsilon, D, k$$

$$P_{\text{stop}} [L_D(\lambda_S) \leq \min_{\lambda' \in \mathcal{L}} L_D(\lambda') + \epsilon] \geq 1 - \delta$$

למשל תחילה נ-  $U_C$  אז  $f_0$  וכן  $m_n(\varepsilon, f) \leq m_n^U(\varepsilon, f)$

$$\frac{\varepsilon}{2} \quad \text{אז} \quad S - p \geq 0, \quad m \geq \ln_{\frac{p}{2}} \left( \frac{\varepsilon}{2}, \gamma \right) \quad \text{אז}$$
$$L_D(h_S) \leq \min_{h \in \mathcal{H}} L_D(h) + \epsilon$$
 is just uniform representitive.

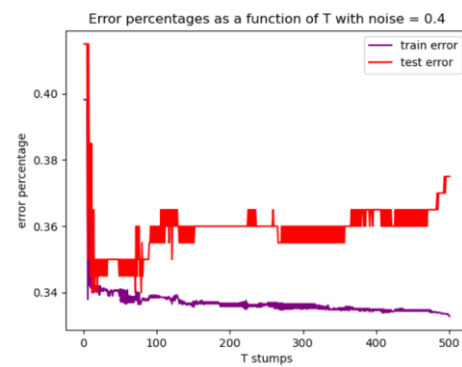
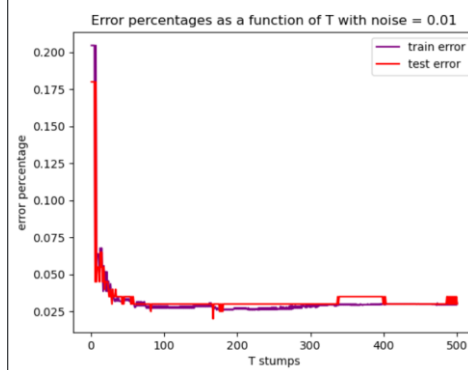
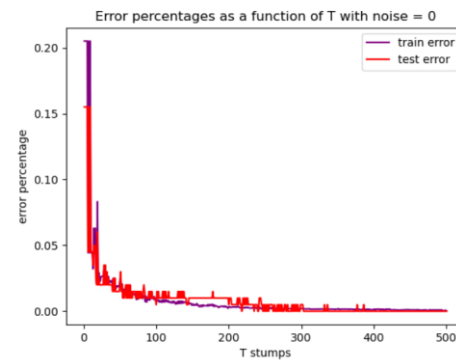
10/11 (10/11) 10/11

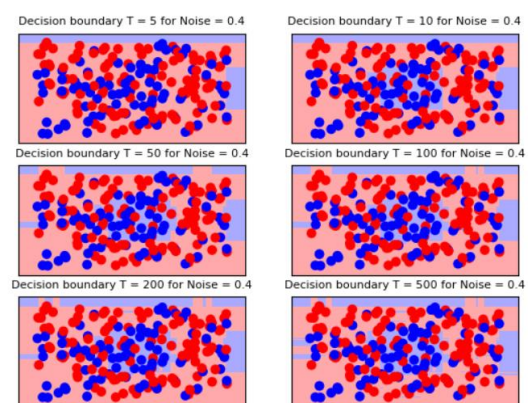
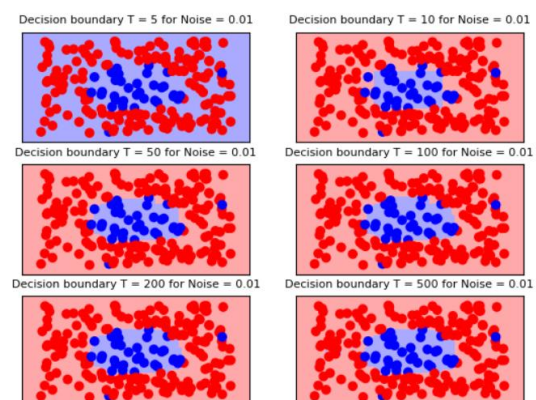
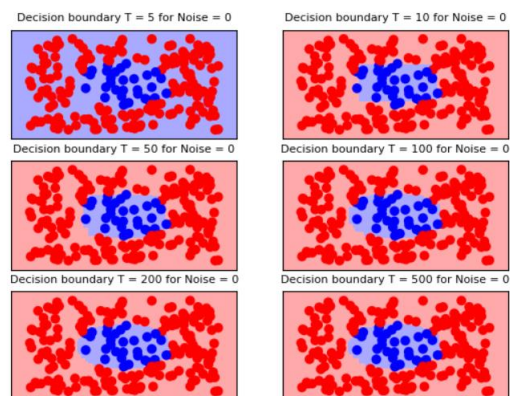
$$P_{\text{sup}} [L_D(h_S) \leq \min_{h \in H} L_D(h) + \varepsilon] \geq D^n(\{S_0(x, y) \mid \begin{smallmatrix} \text{is a } S \\ \text{representative} \end{smallmatrix}\}) \geq 1 - \delta$$

2022



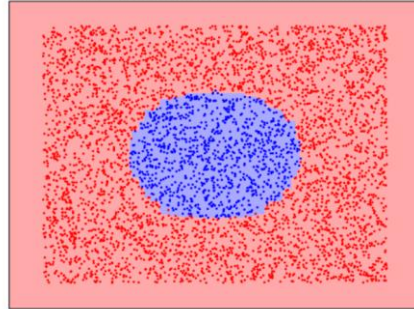




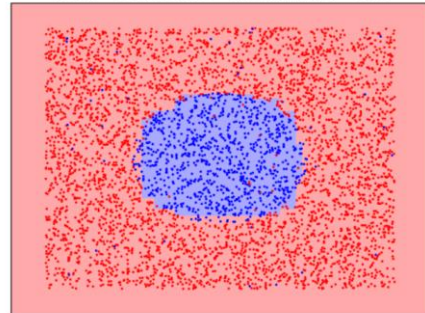




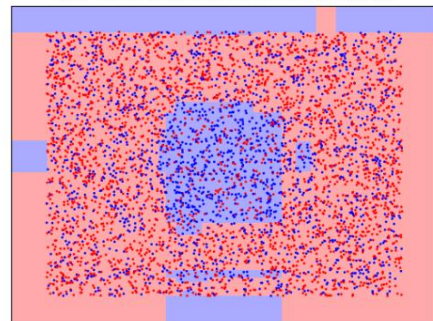
Best T = 218 for Noise = 0 with Error = 0.005

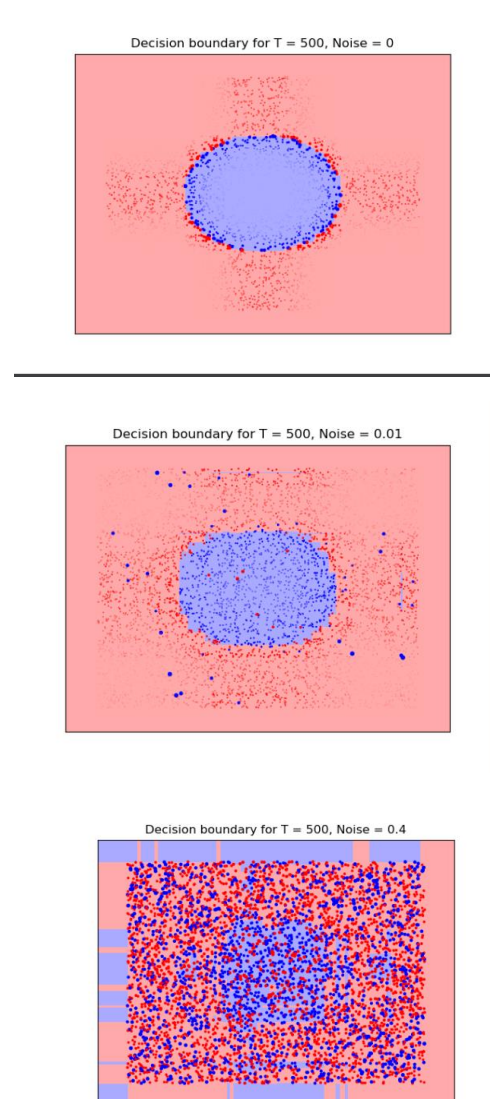


Best T = 167 for Noise = 0.01 with Error = 0.03



Best T = 89 for Noise = 0.4 with Error = 0.385





ניתן לראות שיש השפעה משמעותית של הוספת הרעש, על ההצלחה של 500 מסווגים להצליח לסווג נקודות כחולות\אדומות שמוכלות ממש באזורים אדומים\כחולים (בהתאמה), ולכן ככל שעולה הרעש, נראה יותר נקודות בולטות. נוכל לראות שיש לכך גם השפעה רבה לגבי ה-decision boundary של התוצאה הסופית, בייחוד כשיש רעש של 0.4.

17. בשאלה 13 ניתן להבחין את ההשפעה של הרעש על ה-

bias-variance tradeoff. ככל שהרעש וכמות המסווגים עולה, ניתן לראות שכך גם האחוז שגיאה עולה. לעומת זאת, כאשר אין רעש כלל, ככל שנוסיף יותר מסווגים כך כמות השגיאה יורדת עד סף מסויים. זה גם מתבטא בשאלה 15 בכך שככל שהרעש עולה, כך גם כמות המסווגים האופטימלית יורדת.