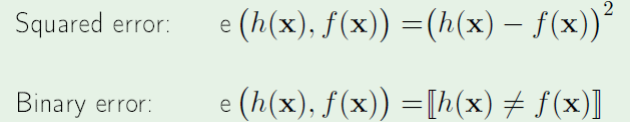
Machine Learning Theory – תאוריית למידה ממוכנת

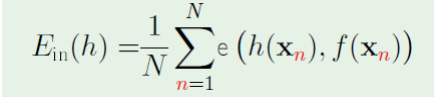
Error measures– צעדי השגיאה

שתי שיטות למדידת הטעות: שגיאה ריבועית

שגיאה בינארית (0/1)

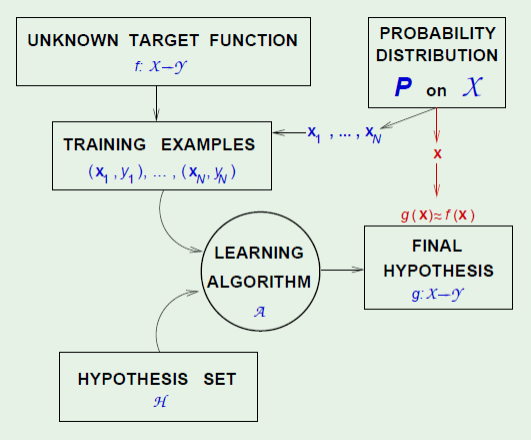
מנקודתיות לכוללניות

שגיאה כוללנית = ממוצע של השגיאות הנקודתיות

שגיאות בתוך המדגם:



שגיאות מחוץ למדגם:

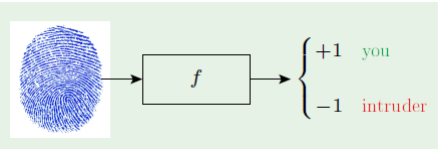
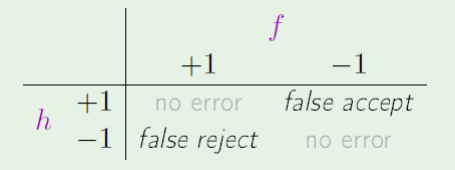


דיאגרמת הלמידה – גרסה 3

למידה חדשה שנוספה למודל היא ההשערה ש

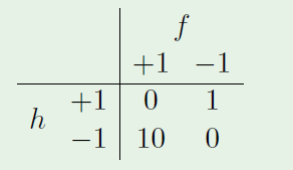
G(x) = f(x)

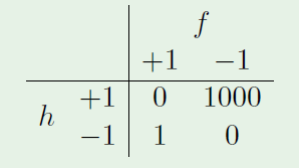
כיצד לבחור את מידת השגיאה?

דוגמא: אימות טביעת אצבע  
ישנם שני סוגים של שגיאה: קבלה לשווא (false accept) או דחייה לשווא (false reject)  
האם הם אותו הדבר?

כאשר ההשערה (h) היא 1+ ובפועל (f) 1- אז זו קבלה לשווא.  
כאשר ההשערה (h) היא 1- ובפועל (f) 1+ אז זו דחייה לשווא.

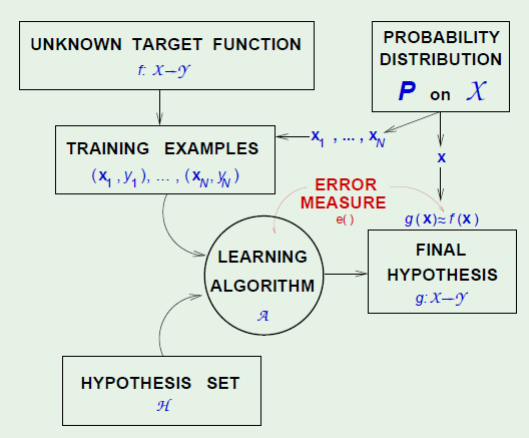
מידת השגיאה – עבור סופרמרקט:  
הסופרמרקט מזהה טביעת אצבע עבור הנחות. דחייה לשווא היא יקרה, לקוחות עלולים להתעצבן על כך שמגיע להם הנחה ולפי פונקציית ההשערה לא מגיעה הנחה.  
קבלה לשווא, זו מידת שגיאה מינורית (קטנה). הסופרמרקט יכול לספוג את השגיאה הזו.



מידת השגיאה – עבור ה CAI  
ה CAI בודק טביעת אצבע בשביל האבטחה. קבלה לשווא היא אסון (הכנסת אדם שאינו שייך ל CAI). ודחייה לשווא יכולה להיות נסבלת, פשוט צריך לנסות שוב (האדם כן שייך)

ניתן לראות שבמידת השגיאה עבור הCAI יש שגיאת הרבה יותר גדולה, מפני שהשגיאה יכולה לגרום לנזק.

ציון מידת השגיאה – מדד השגיאה צריך להיות מוגדר ע"י המשתמש אך לא תמיד אפשרי.   
חלופות נפוצות: squared error ≡ Gaussian noise, closed-form solutions.



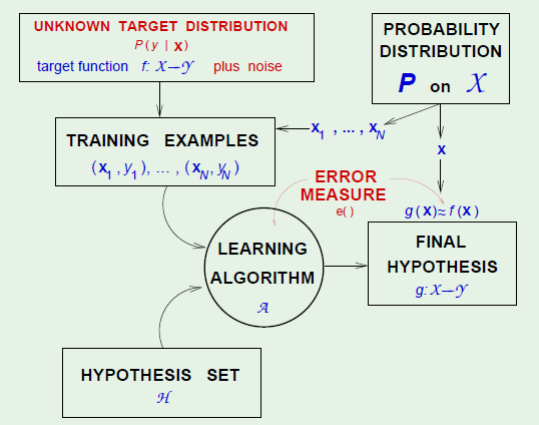
דיאגרמת הלמידה – גרסה 4

דבר חדש שנוסף למודל שיעזור לבנות מודל טוב יותר הוא מידת השגיאה. (באדום)

Noisy target– מטרה רועשת

פונקציית המטרה היא לא תמיד פונקציה. לדוגמא: באישור האשראי. שני לקוחות זהים (מבחינת התכונות) אבל שתי תוצאות שונות.  
יש כאן רעש במטרה. קיימת הסתברות כלשהי לאישור האשראי.

Target distribution– התפלגות המטרה

* במקום y = f(x) אנו משתמשים בהתפלגות המטרה: P(y|x) (בהינתן x מה ההסתברות ש- y יקרה?)
* (x,y) עכשיו נוצרו מההתפלגות המשותפת P(x) P(y|x)
* מטרה רועשת – מטרה דטרמיניסטית f(x) פלוס הרעשים
* מטרה דטרמיניסטית היא מקרה מיוחד של מטרה רועשת (רעש = 0)

דיאגרמת הלמידה – גרסה 5

דבר חדש שנוסף למודל שיעזור לבנות מודל טוב יותר הוא התפלגות המטרה + רעשים

The 2 questions of learning– 2 השאלות של הלמידה

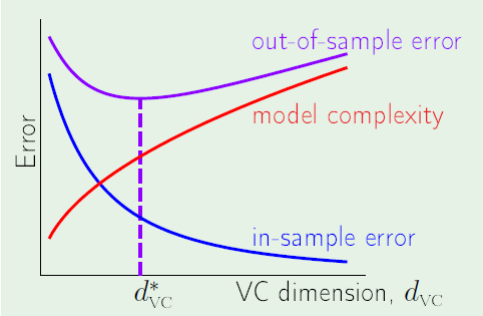
יש מצבים שבהם: Eout(g) ≈Ein(g) (השגיאה בתוך המדגם שווה בערך לשגיאה שמחוץ למדגם). זה נקרא הכללה (**generalization**)  
האם זה למידה? אנו צריכים ש f≈g, מה שאומר ש Eout(g)≈ 0

1. האם אנחנו יכולים לוודא ש Eout(g) הוא קרוב מספיק ל- Ein(g)?  
2. האם אנחנו יכולים לעשות את Ein(g) קטן מספיק?

Complexity of a model– מורכבות של מודל

* עבור מודל מורכב יותר קל להתאים מערך נתונים טוב. אבל מה עם ההכללה?
* בתיאוריה של Vapnik-Chervonenkis’ קיימת הוכחה סטטיסטית של הכללה מלמידה ממוכנת, האם הגודל של מערך הנתונים הוא מספיק גדול.
* לפי VC, ממד הdVC מאפיין את רמת המורכבות של המודל (סט ההשערה)
* dVC≈ מספר הפרמטרים העצמאיים (הבלתי תלויים) ב- h (פונקציית ההשערה שלנו).

האיזון של המורכבות - The tradeoff of complexity



בתרשים רואים שככל שה dvc גדול יותר אז:  
כאשר מורכבות המודל עולה (אדום), כך השגיאה בתוך המדגם קטנה יותר (Ein) וההפרש בין השגיאה בתוך המדגם (Ein) לשגיאה מחוץ למדגם(Eout) גדלה יותר.

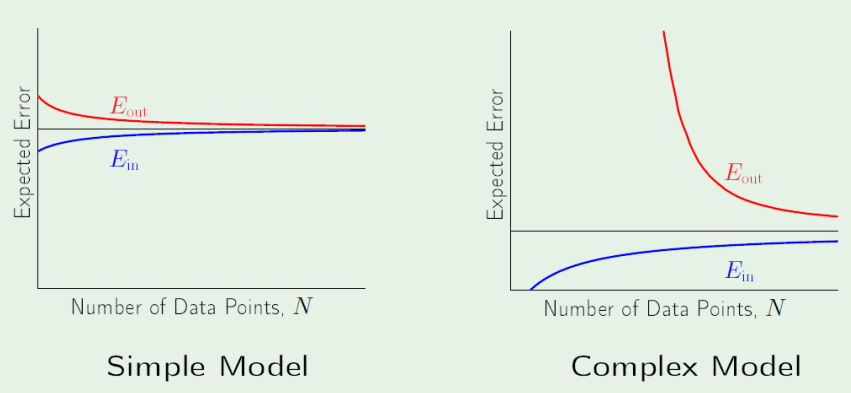
כלל ההכללה - The generalization rule

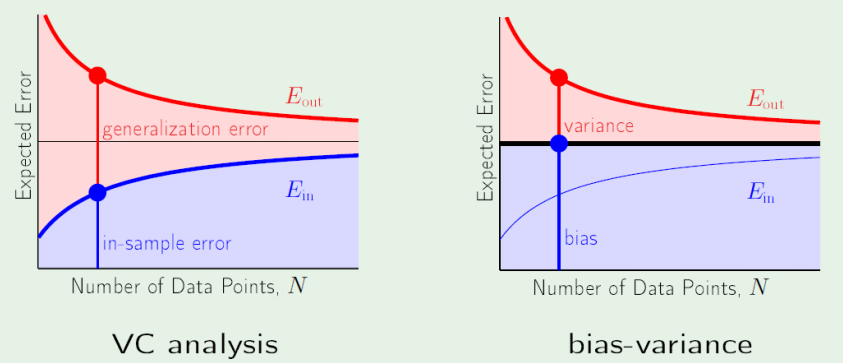
* בשביל מספר ההכללה המקובל של נקודות בסט אימון, N צריך להיות משמעותית גדול מ –Dvc.
* כלל אצבע: dvc 10 ≤N
* למודל מורכב אתה צריך שיהיה מספיק נקודות של נתונים.

הטיה ושונות - Bias and Variance

* איך סט השערות H טוב יכול לשער את f?
* אפילו במקרה של N מאוד גדול יכול להיות שגיאה. שגיאה זו נקראת הטיה של המודל (**bias**). לדוגמא: עם מודל לינארי לא ניתן לשער נתונים לא לינאריים. ההטיה יכולה להיות שווה לאפס רק אם f נמצאת בתוך H.
* שונות (variance) היא חלק נוסף של Eout הקשורה למערך הנתונים (גודל מוגבל ורעש) ולמגבלות של אלגוריתם הלמידה.

למידת העקומות



VC vs. bias-variance

למידת העקומות עבורlinear regression

