

ואם יתבטא בגזאלטריציה.

עוד משהו קטן לפני להאמר ניכנס לדוגמה מנחה 2-אומרים:

בהינתן תצפיות, כלומר מיצו חלקי הנלקח ממוצע כלשהו עם פרמטרים, ניתן לאמוד

את ערך הפרמטרים האמיתיים, נסמנו ב- w , ע"י ערך שלנקה בהתבסס על התצפיות ונסמנו $\hat{w}$.

כמובן שנרצה ל- $\hat{w}$ יהיה קרוב ל- w .

במקרה של ^{שימוש ב-}סטטיסטיקה בייסיאנית, יש אומדן נראה מריבוי $\hat{w} = \arg \max_w P(z|w)$,
(Maximum Likelihood Estimator, MLE)

דוגמה:

במקרה של התפלגות גאוסיאנית עם פרמטרים μ, σ^2 ותצפיות $\{x_i\}_{i=1}^n$ שלנקהו מ- $N(\mu, \sigma^2)$,

עוד הנראות נלקח כמו

$$P(\{x_i\} | \mu, \sigma^2) = \text{const} \cdot n \cdot \sigma^{-n} \cdot \frac{1}{2} \sum_i \frac{(x_i - \mu)^2}{\sigma^2}$$

לפיכך מקסימום נראה ב- μ ו- σ (גזירה לפי μ ו- σ והנראה ל-0)

נומרת אומרים לעריכה:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2$$

האומדן ל- σ^2 מוטעה - היינו מצפים ~~להאומדן של σ^2~~ $E[\hat{w}] = w$ (חסר הי"ס),

אבל במקרה של $\hat{\sigma}^2$ זה לא נכון. התיקון שנעשה עבור $\hat{\sigma}^2$ הוא החלפת n במכנה ל- $n-1$.

~~היחס בין $\hat{\sigma}^2$ ל- σ^2 הוא $\frac{n-1}{n}$ ולכן $\hat{\sigma}^2$ הוא אומדן מוטעה של σ^2 .
כדי לקבל אומדן של σ^2 שיהיה בלתי מוטעה, נשתמש ב- $\frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mu})^2$.~~

הערה: לתיקון $n-1$ קוראים תיקון בסק, (הוא בא כשמעריכים ממוצע ואם סטיית תקן

והאינטואיציה שלו יכולה להגיע מכך שבמקרה של n גזירות יש דרישה חופש אחת

שהנוסחה של מציאת הממוצע, והשארים עם $n-1$ דרישות חופש בלבד להעריכה העולה.

הקדמה של כל מה שדיברנו ועוד:

~~הקדמה~~ הקדמה

נניח ויש לנו דגימות הנלקחות מהמוצד הבא: $y = \sin(2\pi x) + \eta$

כאשר η רעש אקראי שלקח מהתפלגות גאוסית עם תוחלת 0 ושונות לא ידועה.

(אבל באופן כללי η רעש אקראי מהתפלגות כלשהי $y = f(x) + \eta$).

~~הקדמה~~

אם הדגימות נסמך ה- $\{x_1, \dots, x_n\}$ ואם הערכים (תוצאות/targets) נסמך ה- $t_1, \dots, t_n$.

~~הקדמה~~

~~הקדמה~~

נרצה למצוא קבוצת משקל w נלקחו הדגימות x_i הדגימות הנלקחות:

$$y(w, x_i) = \sum_{j=1}^m w_j x_{ij}$$

אז בעצם הפרמטרים שנרצה למצוא הם $\vec{w} = (w_1, \dots, w_m)$.

הדרך: נרצה למצוא פונקציה של w $E(w) = \frac{1}{2} \sum_{i=1}^n (y(w, x_i) - t_i)^2$ (ככה עושים בריבוע).

לעבודה $\{x_1, \dots, x_n\}$ קוראים training set, כי בעצרת אנחנו למדנים את המוצד.

אסור למצוא אפילו אנלטיק את w^* (האופטימלי, עמרו $E(w^*)$ מינימלי).

האם זה אומר שמצאנו פתרון טוב? כן, בהכרח.
 כדי לבדוק אם הפתרון טוב, בדרך עוקבים קבוצת טסט $(test)$ ומדנים את השגיאה $E(w)$ (אנלטיקלי), אם ניקח $m = n-1$, אז אפשר למצוא w^ כך ש- $E(w^*) = 0$ (אנלטיקלי), יחסית אליה.

אבל כמובן שפולטנום ככה לא יתאר היטב את המוצד אלמא יראה משהו ככה:



מצב זה נקרא Overfit - התאמת יתר E_{train} .

כדי למנוע זאת, נרצה להאמין פחות לעבודה האימון - נשים regulization:

$$\tilde{E}(w) = E(w) + \frac{\lambda}{2} \|w\|^2$$

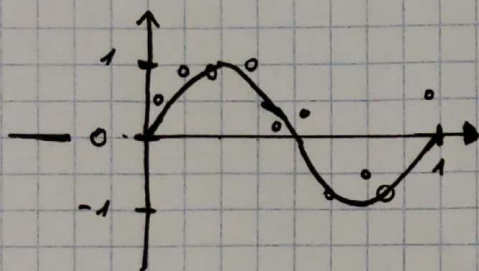
ואכן, זה עוזר למנוע את ה- Overfit. ראה ציורים בהמשך.

(אם כי הוולרייזציה ככה קוראים L_2 או ridge, שהבדל מ- L_1 או LASSO, שהוא סכום הערכים המוחלטים של איברי w . כמובן שיש איברי וולרייזציה אחרים, אך אלה הנפוצים ביותר).

הערה - LASSO שונה מ- ridge בכך שהוא מאפשר רכיבים ספציפיים של w ייתר טוב.

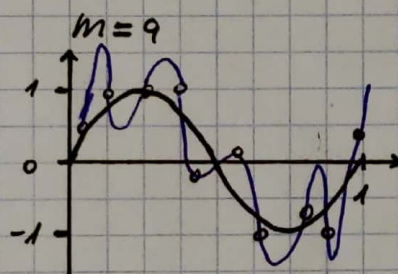
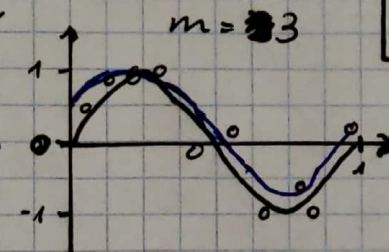
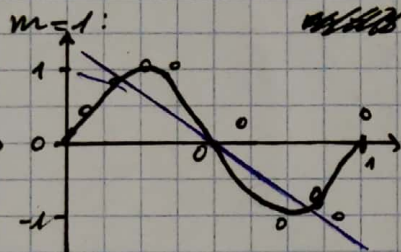
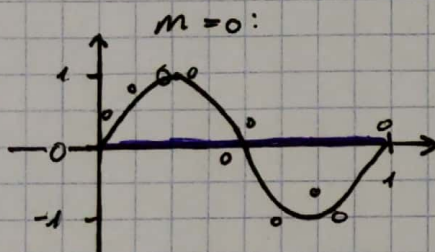
הערה - מצב שני, Underfit, גרע טוב מדי "מן סוף Underfit, פירוט מצב שני אין ברירה אלמא עדיף loss גרוע מאוד.

זמרים : הנצח - $\sim N(0, 0.3)$ כאשר



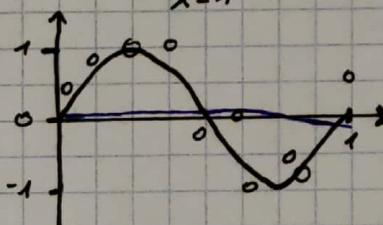
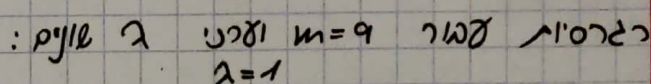
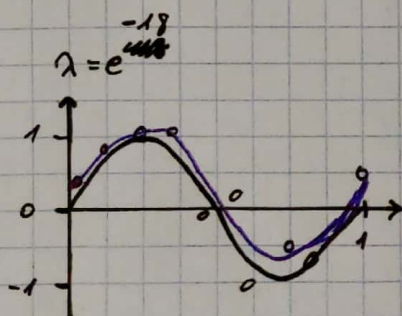
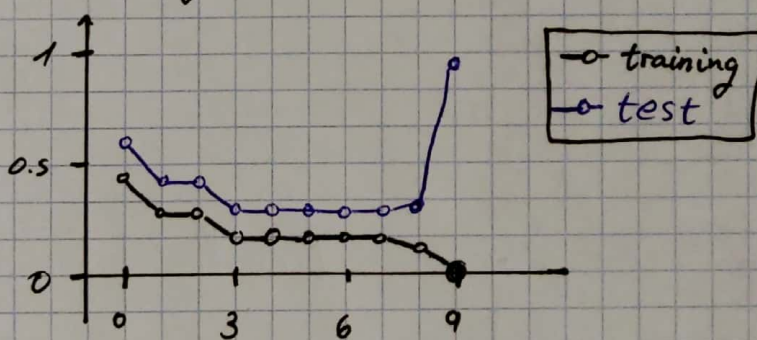
הפועלים הנכונה והזמינה:

רצח ס'ור שנות עשר דגמא פאליצוס שנות: (ללא האלריצ'ה)

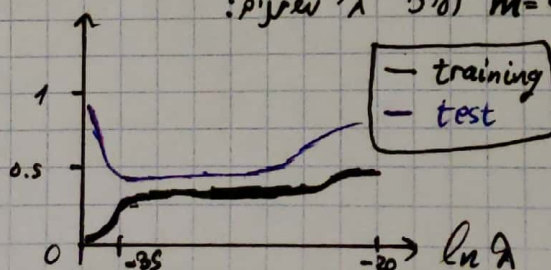


RMS error: $\sqrt{\frac{2E(w^*)}{n}}$

צמוד צרכי מחנכים: (בצורה test - training)



RMS error 748 $m=9$ (צ"ו) λ new pipe:



★ העשרה

75N

Pattern Recognition and Machine Learning

Christopher
M. Bishop