

מבוא למינה מסאכונות - תרבות 1

ע'ת'ים סכ'נים של ת'ות קורס:

מ'ים של - levinai@biu.ac.il

הרכב צ'ון:

100% מבחן

* ט' תר'ים של שבוע, לכל שבוע. נכון לעכ'יו, אין חוב' הישה.

התר'ים יהיו תכנותיים או תיורטיים בהת'ם. חומר שנלמד באופן שבועי. תכנות - python.

נושאי שיעור:

חברה של סטטיסטיקה בייסיאנית, אומ'ים, הסקה סטטיסטי.

- הוק'יה המושגים train-test, פונקצ'יה loss, רגולריזצ'יה, אומ'ן פרמטרים, Prior, Overfit.

סטטיסטיקה בייסיאנית

המדידה שלנו נעשה להתאמה למודל. בהינתן (או בהנחה) שקראנו מסוים נתון מיתר מודל

יבוא עם פרמטרים לא ידועים, נסתמך ב-w, ובהינתן תצפיות של ערכים של נתון

מיתר המודל, נסתמך ב-z, היינו רוצים למצוא את הערך המקסימלי של $P(w|z)$

(או שאר ערך w שלטובנו זה מתקבל).

כפי נוסחה ב'ים,
$$P(w|z) = \frac{P(z|w) P(w)}{P(z)}$$

מה של התקבולו באופן ימין אומרת?

$P(z)$ - לא משהו חשוב במיוחד, זה סתם איזוטר נרמול: $P(z) = \int P(z|w) P(w) dw$

$P(z|w)$ - זו הנראות (likelihood) של המודל - מה היכנוי שקראנו את התצפיות

שקיבלנו בהינתן פרמטרים ידועים.

$P(w)$ - היכנוי הפרמטרי מודל כפי שיתקבלו. ע'ה קוראים Prior, כי זה ע'ז הנחה

~~שנוצרה לפנינו~~ לפני ידערת כל תצפיות מהמודל.

כאשר מ'ים את העסק, משיעים ע'כ שרוצים למקסם פונקצ'יה הנקראת פונקצ'יה

המטרה ~~היא למצוא את הפרמטרים הטובים ביותר~~ (loss).

איתה פונקצ'יה מופנה משני מחוברים עיקריים: פונקצ'יה לג'יאה ואיבר רגולריזצ'יה.

איבר הרגולריזצ'יה הוא ביטוי עכמה אנחנו סומכים על ה-Prior שלנו. למשל, בהיטל

מטבע אשר לקבוע Prior ~~הוא~~ עבור התפלגות היכנוי ע'ים.

אם אנחנו מאמ'נים שהמטבע הולך, נר'בב לקחת  Prior ~

ואם אנחנו פתוח מאמ'נים, נוכל לקחת  . מ'ם מקרה $\sigma^2 = 0.5$ נר'בב לקבוע גם ע'ה כמה נאמין,

ונה יתבטא בגורמים צדדי.

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

בהינתן תצפיות, כלומר מיפץ חלקי. הנלקח ממוצע כלשהו עם פירמטרים, ניתן לאמוד

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

נמאן שנרצה ל- \hat{w} יהיה קרוב ל- w .

במקרה של ^{לימוד} סטטיסטיקה בייסאנית, יש אומדן נראה מירבה, $\hat{w} = \underset{w}{\operatorname{argmax}} P(z|w)$,
(Maximum Likelihood Estimator, MLE)

דוגמה:

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

עוד הנראות תואר כמו

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

לקיחת מקסימום במאמר ב- μ ו- σ (זעיר לפי μ ו- σ והנראה ל-0)

נומרת אומרים עדיכיים:

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

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

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

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

~~האומדן של σ^2 מוטעה כי הוא לא נכון. התיקון שנעשה עבור $\hat{\sigma}^2$ הוא החלפתו במנה $n-1$.
האומדן של μ הוא נכון כי $E[\hat{\mu}] = \mu$.
האומדן של σ^2 מוטעה כי $E[\hat{\sigma}^2] < \sigma^2$.
האומדן של σ^2 מוטעה כי $E[\hat{\sigma}^2] < \sigma^2$.~~

הערה: לתיקון $n-1$ קוראים תיקון בסק, הוא בא כמתמטיים ממוצע ואם סטיית תקן והאינטואיציה שלו יכולה להגיע מכך שבמקרה של n זרימה יש זרימה חופש אחת שהיא לא חופשית הממוצע, והשארים עם $n-1$ זרימות חופש בגבם להערכת העלות.

הקצמה של כל מה שדיברנו ועוד:

~~הקצמה~~ הקצמה

נניח ויש לנו דגימות הנלקחות מהמוצד הבא: $y = \sin(2\pi x) + \eta$
כאשר η רעש אקראי שלקחה מהתפלגות גאוסית עם תוחלת 0 ושונות לא ידועה.
(אבל באופן כללי η רעש אקראי מהתפלגות גאוסית - $y = f(x) + \eta$)

~~הקצמה~~

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

~~הקצמה~~

נרצה למצוא קבוצת משקל ממונן נלקחו הדגימות ע"י הדגימות הנמנות:

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

אם נעצם הפרמטרים שנרצה להעריך הם $\vec{w} = (w_1, \dots, w_m)$

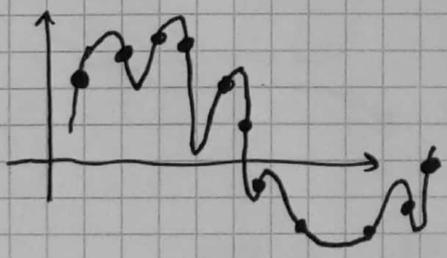
הדרך: נרצה למצוא פונקציה שלישית $E(w) = \frac{1}{2} \sum_{i=1}^n (y(w, x_i) - t_i)^2$ (ככה עושים בריזוסיה)

לקבוצה $\{x_1, \dots, x_n\}$ קוראים training set, כי בעצרתה אנחנו לומדים את המוצד.

אסור למצוא אפילו אנליטי את w^* (הוא פוטנטי), צערו $E(w^*)$ מינימלי.

האם זה אומר שמצאנו הפתרון טוב? או בהכרח. test (למונחים את השיאה) $E(w)$ הוא 0 (אינפיריטיות), אם ניקח $m=1$, אז אפשר למצוא w^* כך של- $E(w^*)$ הוא 0 (אינפיריטיות), יחסית אלה test, אם ניקח $m=1$, אז אפשר למצוא w^* כך של- $E(w^*)$ הוא 0 (אינפיריטיות), יחסית אלה test.

אלה כמובן שפולטנום ככה לא ימארו היטב את המוצד אלמא יראה שזהו ככה:



Overfit - התאמת יתר עם train.

כדי למנוע זאת, נרצה להאמין פחות לקבוצת האימון - נשים regularization:

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

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

(אם ירצה רגולריזציה ככה קוראים L_2 או ridge, שהגזים מ- L_1 או LASSO), שהוא סכום הערכים המוחסטים של איברי w . כמובן שיש איברי רגולריזציה אחרים אך אלה הנפוצים ביותר.

הערה - LASSO שונה מ- ridge בכך שהוא ממצד רכיבים ספציפיים של w יותר טוב מ- ridge.

הערה - מצד שני, ג' גבוה מדי ייתן לנו Underfit, פירושה מצד שני אין הכוונה אלמא שהיות עם loss גבוה מאוד.

