

دستگاه ضعیف اول : با نجوم به آنکه این مدل overfit است

بعن واریانس آن بالا بوده است اما بایان آن می توان لغت ریاضی نموده است.

بنابراین می بایست واریانس آنرا کاهش دهیم. برای این کار ابتدا ^{ترنین} استفاده می نمیسیم. این تکنیک با ساختن weak learner های مستقل از هم و در فن میانلين

از نتایج آنها، باعث می شود واریانس عمل کاهش یابد. جراحت طبیعی میانلين کوچق

از نمونهای مستقل نجوم طبعی واریانس میان آن نمونهها می شود

دستگاه ضعیف دوم : در این دستگاه بزرگ باس بala و واریانس کم است. به همین دلیل

از تکنیک Bagging نسیان استفاده در جراید نه تنها می توانند نجوم طبعی واریانس

شود و ماتری راهی بایامن نمی کنند. به همین دلیل از تکنیک Boosting استفاده

می نمیسیم. این تکنیک در هر iteration سعی می کند (برای) داده هایی را که باید مهمله نماید

که در iteration قبل نتایج ضعیفند (که آنها ذهن است). این کار باعث می شود هر یک از

آنها می تواند نجوم طبعی را به خوبی باید نماید و تکنیک آنها در weak learner

آنها می توانند نجوم طبعی را به خوبی باید نماید و هر یک از آنها می باید نهاد

جراید به محل نسبتاً محدود نمونه های تبدیل شده است

$$\alpha = \frac{1}{2} \log\left(\frac{1 - \epsilon}{\epsilon}\right) \quad \text{اگر خطا ممکن باشد داریم:}$$

$$= \frac{1}{2} \log\left(\frac{1 - 0.5}{0.5}\right) = 0$$

این یعنی آنکه عملیات reweight آتفاق نمی‌افتد و هماند بردا

قبل ادامه در نظر چرا داده AdaBoost این مدل طلا دندم کامل نیست

و چرا دخوا نهاده است یعنی Uniform در نتیجه این یک

نیست و «نتیجه یادگیری بر دسته Boosting اعطان نیز نیست

اگر خفا بیشتر از ۵۰٪ باشد، $\alpha < 0$. یعنی α منفی خواهد بود.

این باعث می‌شود که عبارت $\exp(-\alpha y \hat{y})$ دسته باشیم

یعنی هنگامی که استنباط رخ داده باشد، به جای آنکه وزن آنرا افزایش دهیم، کاهش

فرمیم، یعنی برخلاف نفع اصل AdaBoost، بیشترینها بر علیس می‌شوند

و میزان مثال رهن خواهی ۱۰۰٪ باشد، یعنی مدل دنیا بر علیس عمل نموده است در نتیجه

هر نفع ماهم بر علیس شدن است. بطور کلی AdaBoost فرض می‌نماید که وقت

در آینده وزنها

base learner آن بالای ۷۵٪ است چرا دنیا اینقدر نجات نیاز ندارد

Subject :

Date

$$d^{(0)} = \left[\frac{1}{6} \quad \frac{1}{6} \quad -\frac{1}{6} \right]$$

(ب)

دست زده پایه استخابی می باشد دلایل هست مدل ۱۰۰٪ باشد خط

$$\rightarrow F^{(1)} : y = 1,0 \quad \begin{array}{c} 1,0 \\ \hline 0 \quad 0 \quad 0 \end{array} \times \quad \text{بر استخاب می شم :}$$

$$\hat{\epsilon}^{(1)} = \left(\sum (0 \times \frac{1}{6}) \right) + 1 \times \frac{1}{6} = \frac{1}{6} \Rightarrow \alpha^{(1)} = \frac{1}{2} \log 5 = 0,804$$

$$\Rightarrow d^{(1)} = \frac{1}{Z_1} \left(d^{(0)} \exp(-0,8 y_n \hat{\epsilon}_n) \right)$$

$$= \frac{1}{Z_1} \left(\underbrace{0,07 \quad 0,07 \quad 0,37 \quad 0,07 \quad 0,07}_{0,07} \right) \rightarrow Z_1 = 0,72$$

$$\Rightarrow d^{(1)} = (0,1 \quad 0,1 \quad 0,51 \quad 0,1 \quad 0,1)$$

حال با توجه به آنکه داده سوم روز بینتری در داده مدل به صورت منسوب شد

این داده دست بینی نشود، این بر خط $x = 0,5$ بر نظر نمایم

$$\hat{\epsilon}^{(2)} = \left(\sum 0 \times 0,1 \right) + 0 \times 0,51 + 0,1 = 0,1$$

$$\rightarrow \alpha = \frac{1}{2} \log(9) = 1,098 \approx 1,1$$

$$d^{(2)} = \frac{1}{Z_2} \left(d^{(1)} \exp(-1,1 y_n \hat{\epsilon}_n) \right)$$

$$\Rightarrow Z_2 = 0,58 \Rightarrow d^{(2)} = (0,2 \quad 0,05 \quad 0,27 \quad 0,05 \quad 0,5)$$

Subject :

Date

$$\frac{\partial E}{\partial \alpha} = 0 = -\frac{1}{2} e^{-\alpha/2} \sum w_i^{(t)} + \frac{1}{2} e^{\alpha/2} \sum w_i^{(t)} (\dots)$$

$$\Rightarrow e^{-\alpha/2} \sum w_i^{(t)} = e^{\alpha/2} \sum w_i^{(T)}$$

$$\Rightarrow e^\alpha = \frac{\sum w_i^{(t)}}{\sum w_i^{(T)}} = A$$

$$\alpha = \ln(A) = \ln \left(\frac{\sum w_i^{(t)} I(f(x_i) = y_i)}{\sum w_i^{(t)} I(f(x_i) \neq y_i)} \right)$$

$$= \ln \left(\frac{1 - \frac{\sum w_i^{(t)} I(f(x_i) = y_i)}{\sum w_i^{(t)}}}{\frac{\sum w_i^{(t)} I(f(x_i) \neq y_i)}{\sum w_i^{(t)}}} \right) = \ln \left(\frac{1 - \varepsilon}{\varepsilon} \right)$$

Subject :

Date

مسئلہ ۳. آنچہ افراش میں دختہ ها : آر انسل مدل (Random Forest)

لوگو بائش کہ بایاس مدل کان نیاں یوہ است، یعنی مدل مانند است

بے خوبی دارکن لا مدل نہ مدل ہے اسکا طف سمجھو کی ہے است وہی یعنی

بے خوبی دارکن لا مدل نہ مدل ہے اسکا طف سمجھو کی ہے است وہی یعنی

و سادہ بودا از درستہ نیاں است مدل افراش Hypothesis Space

دھیم تا دختہ های تھیم بتائند مدل های سمجھو کر انجام دند. اس کو

با افراش اسک دختہ ها قابل انجام است، جو کہ چانپر لئے ہو داش افراش اسک

دیک دخت تھیم، صحر بے استغادہ از Feature ہیں لیٹر بری split در

ص شد و استغادہ از Feature ہیں لیٹر صحر بے سمجھو کر نہیں مدل در

درستہ میں بایاس می شد.

افراش تعداد دختہ ها : آر انسل مدل (Random Forest) این بائش در مدل overfit

نمود است و طریق نیا کی دلہ، یعنی base learner مدل بایاس می اسٹے از

اہا تعداد آن کا طف ہے است جو کہ بیالیں کن کا ہمچنان طریق بالاں طریق

بنابراین آر تعداد دخت های مفہوم مستقل را بیشتر نمیم، باعث شدن
که با سیانلین لیری، فاریاس کاهش یابد. این بدان دلیل است که دخت های مفہوم
پایه، به علت تکنیک Bootstrapping نسبت از هم مستقل نسند و از علم
آماری می طنیم سیانلین لیری که از نمره های مستقل باعث کاهش فاریاس می شود.
حال بزری مقایسه دین و بیشتراد: افزایش بحق دخت ها نمی تواند نهان
مستقل فاریاس بالا بگذر چرا که مدل های محیطی لازم ندارد در نتیجه افزایش عقیق
تاثیری ندارد بلکه عکس است بیشتر باعث overfitting نمی شود
همچنین افزایش تعداد دخت ها نمی تواند نهان به مستقل بگایس بالا بگذر چرا که
در نتیجه هر جقدر تعداد را افزایش دهیم، با توجه به آنکه وزن رای هر دخت بسان
به طور مطلق است، در نتیجه آن رای لیری نهایی تفاوتس ایجاد نمی شود و در نتیجه مستقل حل نمی شود

مسئلہ ۳ ابتدی نسیان میں دفعہ واریاسن سے بالائی جستجوی تباہی، دفعہ واریاسن

آن نتیجه‌ها بودن مترادفات: $X_1, X_2, \dots, X_n \rightarrow i.i.d$, $\text{Var}(X_i) = \sigma^2$

$$\text{Var}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} < \sigma^2$$

یعنی دادنیم که درست تخمین سعید است و از Overfit نیست

الgoritم Random Forest ، تعلم بياني بحسب ما ذكرنا في المقدمة

ترسلب (رس) و میانلئی فی لور . د باروئس Boatstrapping و انتشار (فدر)

نیز ساده split ، دسته های تضمین آفرینا سنتکس می شوند

دستیجم صحن نعمتی که در اینجا رایت کردیم با میانلین در حقن باشد

کارهای مذکور در این بخش میتوانند در اینجا معرفی شوند.

عکسی که در آن نتایج ممکن است از این جمله overfitting

ستاری ۴. در این روش wrapper، سعی من لستم زیر جهود ای از نیزهها

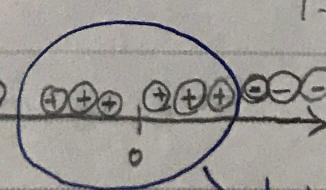
را انتخاب کنید که بهترین دست رفت را برای محیط دارای عرض می‌کند.

دانسته باشند. در تئوری دو دلایل وجود دارد که در overfitting برابر باشد.

مقادیر می‌لند. نخست آنکه فیچرهاي استفاده شوند را متریس دقت بررسی می‌بود Validation طبقه باشد اما خود مدل بررسی طبقه آنرا تأثیرش دیده است. این امر باعث می‌شود که فیچرهاي استفاده شوند در بیشترین قدرات دو تعمیم دهنداشته باشد و نتایج بهتری را دارد. داده‌های که ناجمال نزدیک است ایجاد نند. دلیل اینکه در طور ممکن‌هفظ Feature selection است این است که باعث می‌شود مدل بررسی زیرمجموعه‌ای از دیتای ایجاد شده در نتیجه این تحقیق به لامپری Hypothesis Space می‌شود.

و از ایجاد مدل‌های بسیار ساده جلوگیری می‌کند. مسندی ۵. این جمله علطف است. مثال تففن اول: یکی از معیارهای فیلتر correlation است. یک فیچر را در نظر می‌گیریم که آن با برخی correlation های دیگر مرتبط است اما این نیز استاندارد خوب دسته‌های آن را نمی‌داند.

بدین‌گونه مثال فرض کنیم یک دیتای را استفاده کرد و دارای دو دسته است که

 می‌توانیم تغییر کرده ایم و بین دسته‌ها را از نماینده جدا کنیم: decision boundary

حال بازج بـ \rightarrow Correlation ، معیارهای میزان خلق می باشد، متغیر

است که کوئیش بـ \rightarrow بحسب داده دیش صفر است، هرچهار داده ای

متغیر خلق نسبت داده دیش \rightarrow decision boundary است، لیکن آنها خلق نسبت داده دیش عرضه اند

$$\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} = \text{Pearson Correlation}$$

نتیجه این دیش بـ خوب می توانست داده ها را از هم جدا کند اما کوئیش آن

با بر حسب صفر است

مثال تقاضی \rightarrow $I(X; Y)$ داشت تفہیم داده دیش Mutual Information

لادنفر میگیریم. منظیم داشت تفہیم یک الگوریتم Greedy است که

بـ هر مرحله معیار با بینتریز Mutual Information را انتخاب می کند اما از نوع

بهترین داشت را تولید نمی کند. این بـ \rightarrow دلیل است که این معیار تنوع داده

کیفیت دسترسی نسبت دارند داشت تولید شد از آن زیر فعالیت بهبود فی بود.

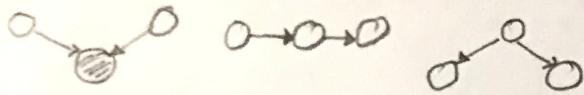
مسئله داشت که \rightarrow ۲ را در نظر میگیریم و مارک بینلت آن را با $B(Y|X)$

نشان میگیم. باقی محرومها در شبکه ای داشتیم، برای آنکه

حالات فعال:

Date _____

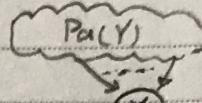
Subject :



اکثر ۲، یعنی بری آنکه وابسته نشوف باشد، می باشد مدل سیرین $A \rightarrow B(Y)$

ازین اتفاقی S به M موجود باشد. نشان می دهیم که همین سیری

سروور نیست.

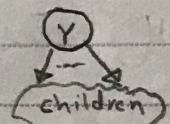


سیر فعال می تواند شامل یال میان ۲ و پرنس باشد

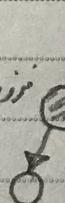
جواب داریم: با توجه به سه مالی که بری فعال

بود در بالا رسم کرده ایم \Rightarrow این دو مردم هر دو غیرفعال هستند

سیر فعال لحی تواند شامل های میان ۷ و فرزندانش

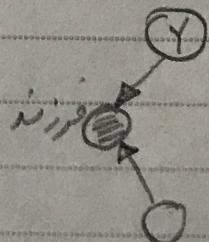


باشد جواب داریم: ۱) فرزند در این حالت با توجه به هائیور خودن



فرزند شرط وسط، سیر غیرفعال است

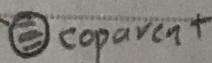
دقت کنید میان دالت به ظاهر فهم به سیر فعال می شود



اما پدر دوم فرزند، در واقع همان coparent است که



می باشد هائیور بخوبی. یعنی سیر نیز غیرفعال



است. دنتیم که سیرها به اند غیرفعال نیستند و یعنی $d\text{-separated}$

\bullet MICRO \rightarrow separated

مسئلہ ۷. نایاب فریبی Lasso بحث است اس کے بعد داد دکھا feature تعداد

$$\text{Cost}(w) = \sum_{i=1}^M (y_i - \sum_{j=0}^N w_j x_{ij})^2 + \lambda \sum_{j=0}^N |w_j|$$

نرم دار w

دینوش، متفاوت است که از نرم دار w

د آن استفاده شده است. این امر باعث شده است که حل مسئلہ

سبک سازی $\min_w \text{Cost}(w)$ منجر به جواب شود که آن جواب sparse

است. یعنی آن جواب w ، طالی مقادیر باری عطف صفر است. این

وضوح بصورت feature selection، باعث می شود embedded

نهان آنفراخ بیفتہ جو کہ بسیاری از عناصر موجود در w صفر شده اند و در

نتیجہ نتیجہ استخراج دیگری های کم می شوند. چنین این منجر به کمی regularizer

نتیجہ overfitting نیز می تواند شود. همین دلیل دارد w حال sparse است

و همین دلیل دلیل اضافه کردن باعث می شود regularizer

$$\text{Cost}(w) = \underbrace{\sum_{i=1}^M (y_i - \hat{y}_i)^2}_{A(w)} + \lambda \underbrace{\sum_{j=1}^N |w_j|}_{B(w)}$$

نیز می شود. جو کردیم.

خطاب با تغییر ضرایب لامدا نوییم:

$$\min A(w)$$

$$\text{s.t. } B(w) \leq u$$

MICRO Overfitting Hypothesis Space دلیل نتیجہ کافش نتیجہ

دروش Ridge - Regression دارم:

$$\text{Cost}(w) = \frac{\sum (y_i - \sum_{j=0}^n w_j x_{ij})^2}{A(w)} + \lambda \sum_{j=0}^n w_j^2 \rightarrow \text{نرم - ۲}$$

این روشن مجهنان با توجه به کاهش overfitting Hypothesis Space

می شود . اما این روشن نه تنها مانند sparse Lasso فتحر به سدن

جواب مستعد بهینه سازی شونده دلیل این تفاوت را اگر به صورت غنیمتی نگرفتیم

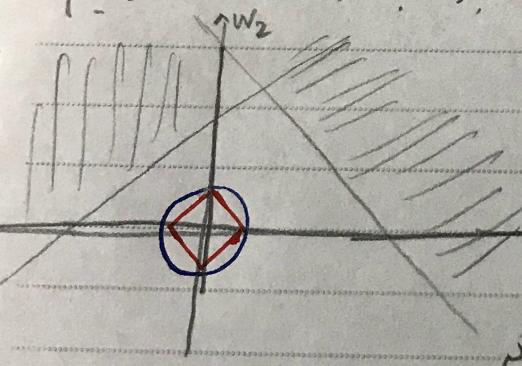
بپسیم : با توجه به حقیقت ضرایب لامبرت می توانیم مستعد بین صورت بازدیدی نیز نیم:

$$\min B(w)$$

$$\text{s.t. } A(w) \leq u$$

این محدودیت constraint

یک محدودیت کانزس است ، در فناوری دو بعدی به صورت خط همان معادله Mean Squared Error



نایم های این محدودیت ، یعنی جایی که تیود برقرار است

حال با توجه به آنکه من فراهم $B(w)$ نمیخواهم

نمیخواهم $\|w\|_1$ و $\|w\|_2$ را کم کنم

زیاد نمیخواهم تا این جایی که تیود برقرار شود . به عبارت سال

$$\rightarrow \begin{cases} 1\text{-norm: } |w_1| + |w_2| = 0,1 \\ 2\text{-norm: } w_1^2 + w_2^2 = 0,1 \end{cases}$$

$$\rightarrow \text{دایره}$$

این روش نایمی شود نرم - ۱ داشت اگرچه قید دلتا نمایند نرم - ۲

این صورت نیست . در نتیجه نرم - ۱ نایمی معتبر شد بلکه از متاصر w حیث روش این نرم - ۲ نمیباشد