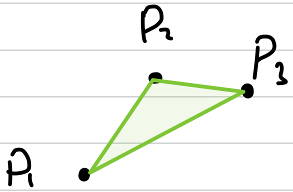
Alireza Habibzadeh 99109393 ML Problem set 2

1. 1) ابتدا نشان می دهیم perceptron در $\mathbb{R}^3$ چهار نقطه را می‌تواند جدا کند.

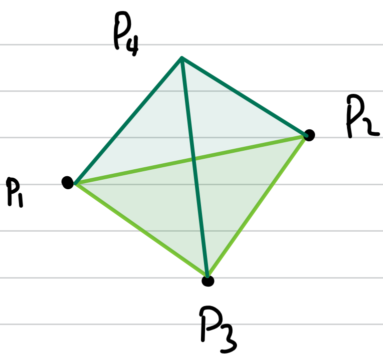
چهار نقطه‌ی دلخواه در $\mathbb{R}^3$ در نظر بگیرید.



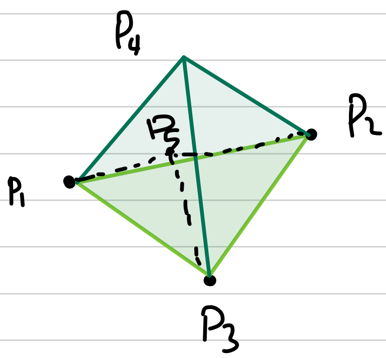
درصورتی که سه تا را یک رنگ و دیگری را رنگ دیگر کنیم درصورتی که نقطه با رنگ متفاوت داخل صفحه‌ی $P_1P_2P_3$ نباشد با یک صفحه به سادگی از بقیه قابل تفکیک است.

پس $k_{min} > 4$

حال نشان می دهیم $k_{min} = 5$. درصورتی که هر نقطه داخل صفحه‌ی سه تای دیگر باشد به طور بدیهی داده ها قابل تفکیک نیستند چرا که می‌توان رنگ آن را متفاوت با سه تای دیگر گرفت و حکم اثبات می‌شود. اما اگر چنین نباشد:



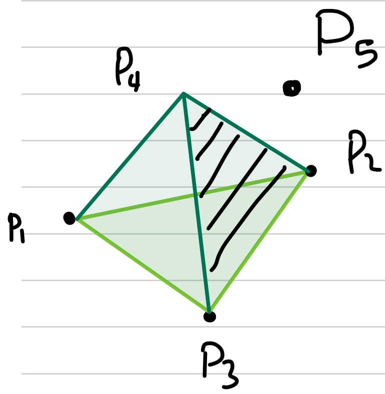
نقاط $P_1, \ldots, P_4$ یک هرم تشکیل می‌دهند. اگر نقطه‌ی P_5 خارج هرم بود ادامه می دهیم. اگر P_5 داخل هرم بود P_5 با سه نقطه‌ی دیگر یک هرم داخلی تشکیل می‌دهد که این هرم قطعاً داخل هرم بزرگ است و نقطه‌ی دیگر (اینجا P_4) بیرون می‌ماند. نام P_4 و P_5 را عوض می‌کنیم



$$P_4 \longleftrightarrow P_5$$

تا P_5 نقطه‌ای باشد که بیرون افتاده .

صفحاتی که $P_1, \dots, P_4$ تشکیل می‌دهند را در نظر بگیرید. هر کدام از این صفحات فضا را به دو قسمت تقسیم می‌کنند. سمتی که رو به هرم است را سمت هرمی و سمت دیگر را سمت خارجی می‌نامیم. اشتراک سمت‌های هرمی 4 صفحه خود هرم است پس P_5 که بیرون هرم است قطعا در سمت خارجی یکی کدام از صفحات قرار دارد.

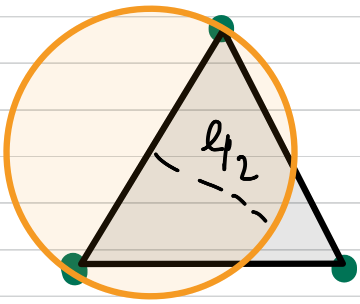


فرض کنید این صفحه $P_2P_3P_4$ باشد.

حال کافی است P_2 و P_3 و P_4 را $+1$ کنیم و P_1 و P_5 را -1 .

هر هایپرپلنی که $P_1, \dots, P_4$ را درست تشخیص دهد قطعا P_5 را هم علامت با P_2 و P_3 و P_4 تشخیص می‌دهد که در مثال ما این طور نیست و 5 نقطه یک نقطه‌ی شکست است و 4 نبود پس 5 کوچک‌ترین نقطه‌ی شکست هایپرپلن در $\mathbb{R}^3$ است.

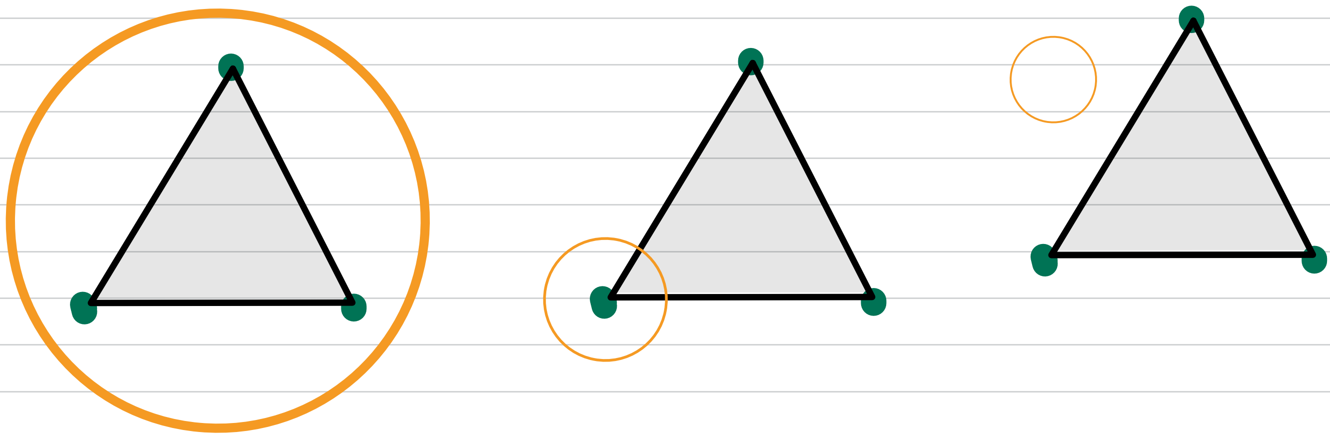
1) سه نقطه را در نظر بگیرید که تشکیل مثلث متساوی الاضلاع 2)
میدهند. دایره‌ای به مرکز وسط ضلع دو نقطه و به شعاع



نصف ضلع هر دو نقطه‌ی طرفین را در بر گیرد ولی
نقطه‌ی سوم را در بر نمی‌گیرد چرا که ارتفاع مثلث
متساوی الاضلاع از نصف ضلعش بزرگتر است :

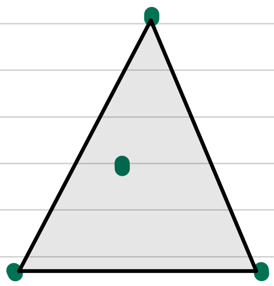
$$l \cos\frac{\pi}{6} = l\frac{\sqrt{3}}{2} > \frac{l}{2}$$

ترکیب‌های سه تایی و یکی‌ای و هیچ‌کدام هم به راحتی قابل تفکیک است :



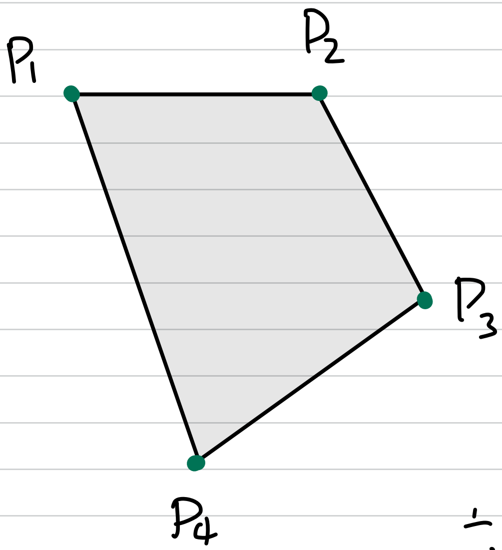
پس سه نقطه را shatter می‌کند و $d_{vc} \geq 3$ و حال نشان می‌دهیم
هیچ قرارگیری 4 تایی را shatter نمی‌کند.
4 نقطه‌ی دلخواه در صفحه در نظر بگیرید. اگر تشکیل یک چهارضلعی

حدب نبودند یک نقطه‌ی داخلی وجود دارد:



دایره‌ای که سه نقطه‌ی بیرونی را احاطه کند حتما نقطه‌ی درونی هم احاطه خواهد کرد پس shatter نمی‌شود در تمام.

اما اگر تشکیل 4 ضلعی محدب بدهند:



حالتِ $P_1, P_3 = +1$ و

$P_2, P_4 = -1$

قابل shatter کردن نیست.

فرض خلف کنیم که چنین دایره‌ای وجود داشته باشد.

از بین دایره‌ها آنی که P_1 و P_3 روی مرز قرار دارند را انتخاب می‌کنیم.

زاویه‌ی $\angle P_2P_3P_4$ زاویه‌ی بین P_3 به دو نقطه خارج دایره است و

$\angle P_2P_1P_4$ هم زاویه بین P_1 به همان نقاط. اگر P_2, P_4 روی مرز دایره بودند

$\angle P_2P_3P_4 + \angle P_2P_1P_4 = \pi$ اما این دو خارج دایره‌اند پس

$\angle P_2P_3P_4 + \angle P_2P_1P_4 > \pi$. اما این دقیقا خلاف محدب بودن

چهارضلعی است پس به تناقض خوردیم و shatter نمی‌شود و $\boxed{d_{VC} = 3}$

ب) ابتدا نشان می‌دهیم در $\mathbb{R}^d$ یک چینش از $2d$ نقطه وجود دارد

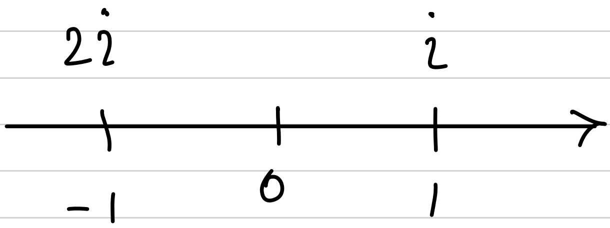
که هر دایکاتومی از آن قابل shatter کردن است.

$2d$ نقطه را به این صورت می‌چینیم که نقاط i و $2i$ مختصات

همه‌ی ابعادشان صفر است به جز بعد i ام که یکی را بر $+1$ و یکی

را بر -1 می‌گذاریم.

حال یک ترکیب دلخواه را در نظر بگیرید. تنها بعد i ام را نگاه می‌کنیم.



اگر می‌خواستیم نقطه‌ی i ام داخل مستطیل d بعدی باشد کافی است مرز مثبت

بعد i ام مستطیل را بزرگتر از 1 قرار دهیم و در غیر این صورت آن را بین

0 و 1 قرار می‌دهیم. برای نقطه‌ی $2i$ هم مرز منفی مستطیل را بین

-1 تا 0 یا کمتر از -1 می‌گذاریم پس آزادی داریم تا هر کدام از

نقاط را داخل یا وارد مستطیل کنیم پس $d_{VC} \geq 2d$

و $\boxed{d_{VC} = 2d}$

۱) نکته‌ی اول این که هر x_i می‌تواند h باشد یا نباشد (درجع) پس در نهایت

2^N تابع h مختلف داریم. اما برای هر دیتای ورودی از x ها به یک h جواب

نیاز داریم پس نهایتا N تا x را می‌توانیم shatter کنیم. پس

$d_{vc} \leq N$

حالا یک مثال می‌زنیم که N تا x shatter می‌شود. کافی است x_i ها

را متمایز قرار دهیم مثلا $x_i = i$ در این صورت h ها انگار

همه‌ی ارقام همه‌ی نمایش‌های باینری هر جمعی از x ها هستند. به ازای

هر ترکیب از خروجی می‌توان h ای پیدا کرد که رقم مناسب را بدهد و

shatter می‌شود. پس $\boxed{d_{vc} = N}$

ج) سوال اول این که آیا تابع خطی f می‌تواند ثابت هم داشته باشد یا نه

اگر می‌تواند f ها به تنهایی مجموعه‌ی صفحات d بعدی هستند

که لزوما مبدا را رد نمی‌کنند. اگر هم خود خطی (ثابت یا

صفر) هم قبول است، باشد حاصل از هم خطی و ... می‌شود و

d_{vc} همچون قبل $d+1$ است.

ادعای من این است که $\boxed{d_{NC}=d}$

اثبات: فرض کنید g در نقطه‌ی $\vec{x_0}$ مقدار g را داشته باشد.

توابع خطی اسکیلر ندارند. می‌توانیم با انتخاب یک پارامتر در تابع خطی f

آن را طوری کنیم که $(f+g)(\vec{x_0})=0$ حال می‌توان d نقطه حول نزدیک

$\vec{x_0}$ طوری انتخاب کرد که g اصلا صفر نباشد و $f+g$ اسکیلر باشد.

در واقع اگر g تابع نرم و پیوسته‌ای باشد این ممکن است. سایر

پارامترهای f خرج برای نقاط کمکی شده و مسئله جواب صفحه‌ی مبدأ گذر است

که یک پارامتر کم دارد.

2.

1. $\epsilon(h_0) \leq \epsilon(h^*) + \eta$ فرض η-optimal بودن

$\hat{\epsilon}(h_0) \leq \epsilon(h_0) + \gamma$ $\left[\gamma_0 = \sqrt{\frac{1}{2m}\log\frac{2k}{\delta}}\right]$

(باند هافدینگ / به احتمال $1-\delta$ چنین است)

$\Longrightarrow \hat{\epsilon}(h_0) \leq \hat{\epsilon}(h) + \eta + \gamma$ (احتمال $1-\delta$)

$\Longrightarrow$ $\boxed{\hat{\epsilon}(h_0) > \hat{\epsilon}(h) + \eta + 2\gamma \qquad \text{احتمال حداکثر } \delta}$

یعنی برای h_0 های η-optimal و δ های کوچک به احتمال زیاد YES داریم.

2. به نظرم حداقل ها حداکثر ها روی بازه هاست. حداقل حداکثر امکان ندارم کلیه باند ها فودینگ دقیقا به همان احتمال آن نامساوی است.

از سمت دیگر باند ها فودینگ استفاده می‌کنیم:

$|\hat{\epsilon}(h_0) - \epsilon(h_0)| \leq \gamma$ به احتمال $1-\delta$

$\Longrightarrow \hat{\epsilon}(h_0) \geq \epsilon(h_0) - \gamma$ ~

(نبودن η-optimal) $\hat{\epsilon}(h_0) \geq \hat{\epsilon}(h^*) + \eta - \gamma$

$$\Rightarrow \quad \hat{\epsilon}(h_0) \geq \hat{\epsilon}(h^*) + \eta - 2\gamma$$

$$\Rightarrow \hat{\epsilon}(h_0) \geq \hat{\epsilon}(\hat{h}) + \eta - 2\gamma \qquad \mathbb{P} = 1-\delta$$

$$\boxed{\Rightarrow \hat{\epsilon}(h_0) < \hat{\epsilon}(\hat{h}) + \eta - 2\gamma \qquad \mathbb{P} = \delta}$$

3. $\hat{\epsilon}(h_0) \leq \epsilon(h_0) + \gamma = \epsilon(h^*) + \gamma \leq \epsilon(\hat{h}) + \gamma$

$$\leq \hat{\epsilon}(\hat{h}) + 2\gamma$$

درحد $m \to \infty$ ، $\gamma \to 0$ میل میکند :

$$\lim_{m\to\infty} \gamma = 0$$

با توجه به تعریف حد اپسیلون و دلتا، حتما m_0 ای وجود دارد که در

$m > m_0$ ، $\gamma < \gamma_0$ میشود. پس میتوان جایی رفت که

$2\gamma < \eta - 2\gamma$ $\left(\gamma < \frac{\eta}{4}\right)$ بشود. پس :

$$\hat{\epsilon}(h_0) \leq \hat{\epsilon}(\hat{h}) + 2\gamma \leq \hat{\epsilon}(\hat{h}) + \eta - 2\gamma$$

که این همان شرط YES شدن است. و $\mathbb{P} = 1-\delta$ و باز به

نظر من احتمال حداقل $1-\delta$ نیست، احتمال خود $1-\delta$ است.

3.

الف)

$$Var(Z_i) = \phi(1-\phi) = \phi - \phi^2 = \underbrace{\frac{1}{4} - \underbrace{(\phi - 1/2)^2}_{\geq 0}}_{\leq 1/4}$$

$\max(Var(z_i)) = \frac{1}{4}$ است که در حالت $\phi = \frac{1}{2}$ اتفاق می‌افتد.

ب)

۱. بدترین ترتیب ممکن چینش جمع احتمال‌ها را برای union می‌نویسیم:

$$\mathbb{P}(|\hat{\epsilon}(\hat{h}) - \epsilon(\hat{h})| > \gamma) \leq \sum \mathbb{P}(|\hat{\epsilon}(h_i) - \epsilon(h_i)| > \gamma)$$

$$\leq \sum \frac{Var(Z_i)}{m\gamma^2} \leq \sum_{1}^{k} \frac{1}{4m\gamma^2} = \boxed{\frac{k}{4m\gamma^2}}$$

که k تعداد h ها است.

اگر کران بالای‌ها از δ کمتر مساوی باشد خود $\mathbb{P}$ هم حتماً کمتر خواهد بود:

$$\frac{k}{4m\gamma^2} \leq \delta \implies \boxed{m \geq \frac{k}{4\delta\gamma^2}}$$

3. در هافدینگ برای ثابت ماندن احتمال باید اندازه نمونه با $\log k$ متناسب می‌بود. ولی اینجا با خود k که رشد میابد، سریع‌تر است.

برای مثال درحالت $k \propto 2^N$ ، $m \propto 2^N$ می‌شود!

ب) مشابه سوال قبل:

$$\epsilon(\hat{h}) \leq \epsilon(h^*) + 2\gamma$$

برای یک احتمال مشخص δ

$$\gamma = \sqrt{\frac{k}{4m\delta}}$$

$$\Rightarrow |\epsilon(\hat{h}) - \epsilon(h^*)| \leq 2\sqrt{\frac{k}{4m\delta}} = \sqrt{\frac{k}{m\delta}}$$

$$\mathbb{P} = 1-\delta$$