University of Tehran
School of Electrical and Computer Engineering

MACHINE LEARNING

Assignment 2

Sasan Keshavarz
810199253

Spring 2022

فهرست

سوال ۱ و ۲ و ۳

-1 $\quad p(x) \sim N(\mu, \sigma^2) \qquad \varphi(x) \sim N(0, 1) \to \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$

$$p_n(x) = \frac{1}{n h_n} \sum_{i=1}^{n} \varphi\left(\frac{x - x_i}{h_n}\right)$$

الف)

$$\bar{p}_n(x) = E(p_n(x)) = \frac{1}{n h_n} \sum_{i=1}^{n} E\left(\varphi\left(\frac{x - x_i}{h_n}\right)\right) = \frac{1}{h_n} \int_{-\infty}^{\infty} \varphi\left(\frac{x-u}{h_n}\right) p(u)\, du$$

$$\bar{p}_n(x) = \frac{1}{h_n} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-1}{2}\left(\frac{x-u}{h_n}\right)^2\right) \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(\frac{-1}{2}\left(\frac{u-\mu}{\sigma}\right)^2\right) du$$

$$= \frac{1}{2\pi h_n \sigma} \exp\left(\frac{-1}{2}\left(\frac{x^2}{h_n^2} + \frac{\mu^2}{\sigma^2}\right)\right) \int_{-\infty}^{\infty} \exp\left(\frac{-1}{2}\left(\frac{u^2}{h_n^2} + \frac{u^2}{\sigma^2}\right) - 2u\left(\frac{x}{h_n^2} + \frac{\mu}{\sigma^2}\right)\right) du$$

$$= \frac{1}{2\pi h_n \sigma} \exp\left(\frac{-1}{2}\left(\frac{x^2}{h_n^2} + \frac{\mu^2}{\sigma^2}\right) + \frac{1}{2}\frac{k^2}{\theta^2}\right) \int_{-\infty}^{\infty} \exp\left(\frac{-1}{2}\left(\frac{u-k}{\theta}\right)^2\right) du$$

$$\theta = \frac{h_n^2 \sigma^2}{h_n^2 + \sigma^2}, \quad k = \theta^2\left(\frac{x}{h_n^2} + \frac{\mu}{\sigma^2}\right)$$

$$\bar{p}_n(x) = \frac{\sqrt{2\pi}\,\theta}{2\pi h_n \sigma} \exp\left(\frac{-1}{2}\left(\frac{x^2}{h_n^2} + \frac{\mu^2}{\sigma^2}\right) + \frac{1}{2}\frac{k^2}{\theta^2}\right)$$

$$= \frac{h_n \sigma}{\sqrt{2\pi}\, h_n \sigma \sqrt{h_n^2 + \sigma^2}} \exp\left(\frac{-1}{2}\left(\frac{x^2}{h_n^2} + \frac{\mu^2}{\sigma^2} - \frac{k^2}{\theta^2}\right)\right)$$

$$\frac{x^2}{h_n^2} + \frac{\mu^2}{\sigma^2} - \frac{k^2}{\theta^2} = \frac{x^2 h_n^2}{(h_n^2+\sigma^2) h_n^2} + \frac{\mu^2 \sigma^2}{(h_n^2+\sigma^2)\sigma^2} - \frac{2x\mu}{h_n^2+\sigma^2} = \frac{(x-\mu)^2}{h_n^2+\sigma^2}$$

$$\longrightarrow \bar{p}_n(x) = \frac{1}{\sqrt{2\pi}\sqrt{h_n^2+\sigma^2}} \exp\left(\frac{-1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right) \longrightarrow \bar{p}_n(x) \sim N(\mu, h_n^2+\sigma^2)$$

$$P_n(x) - \tilde{P}_n(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(\frac{-1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) - \frac{1}{\sqrt{2\pi}\sqrt{h_n^2+\sigma^2}}\exp\left(\frac{-1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right)$$ (ب -1

$$= \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(\frac{-1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right)\left(1-\frac{\sigma}{\sqrt{h_n^2+\sigma^2}}\exp\left(\frac{-1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}+\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)\right)$$

$$= P(x)\left(1-\frac{1}{\sqrt{1+(h_n/\sigma)^2}}\exp\left(-\frac{(x-\mu)^2}{2}\left(\frac{1}{h_n^2+\sigma^2}-\frac{1}{\sigma^2}\right)\right)\right)$$

$$= P(x)\left(1-\frac{1}{\sqrt{1+(h_n/\sigma)^2}}\exp\left(\frac{1}{2}\frac{h_n^2(x-\mu)^2}{h_n^2+\sigma^2}\right)\right), \quad \lim_{h_n\to 0}\frac{1}{\sqrt{1+(\frac{h_n}{\sigma})^2}} \simeq 1-\frac{1}{2}(h_n/\sigma)^2$$

$$\lim_{h_n\to 0}\exp\left(\frac{h_n^2}{2\sigma^2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right) \simeq 1+\frac{h_n^2}{2\sigma^2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}$$

$$P(x)-\tilde{P}_n(x) \simeq P(x)\left(1-\left(1-\frac{1}{2}\left(\frac{h_n}{\sigma}\right)^2\right)\left(1+\frac{h_n^2}{2\sigma^2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right)\right)$$ از h_n^4 به بالا را حذف می‌کنیم

$$= P(x)\left(1-1+\frac{h_n^2}{2\sigma^2}-\frac{h_n^2}{2\sigma^2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right) = P(x)\left(\frac{1}{2}\left(\frac{h_n}{\sigma}\right)^2\left(1-\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right)\right)$$

$$\longrightarrow P(x)-\tilde{P}_n(x) \simeq \frac{1}{2}\left(\frac{h_n}{\sigma}\right)^2\left(1-\left(\frac{x-\mu}{\sigma}\right)^2\right)P(x)$$

$$Var(P_n(x)) = Var\left(\frac{1}{nh_n}\sum_{i=1}^{n}\varphi\left(\frac{x-x_i}{h_n}\right)\right) = \frac{1}{n^2h_n^2}\sum_{i=1}^{n}Var\left(\varphi\left(\frac{x-x_i}{h_n}\right)\right)$$ (ج -1

$\boxed{x_i \to i.i.d}$

$$Var(P_n(x)) = \frac{1}{nh_n^2}Var\left(\varphi\left(\frac{x-a}{h_n}\right)\right) = \frac{1}{nh_n^2}\left(\underbrace{E\left(\varphi^2\left(\frac{x-a}{h_n}\right)\right)}_{I} - \underbrace{E^2\left(\varphi\left(\frac{x-a}{h_n}\right)\right)}_{II}\right)$$

$$I: E\left(\varphi^2\left(\frac{x-u}{h_n}\right)\right) = \int_{-\infty}^{\infty}\varphi^2\left(\frac{x-u}{h_n}\right)P(u)\,du =$$

$$= \int_{-\infty}^{\infty}\frac{1}{2\pi}\exp\left(-\left(\frac{x-u}{h_n}\right)^2\right)\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(\frac{-1}{2}\left(\frac{u-\mu}{\sigma}\right)^2\right)du$$

ادامه در صفحه بعد

ادامه ۱- ج)

$$E\left(\varphi^2\left(\frac{x-u}{h_n}\right)\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-1}{2}\left(\frac{x-u}{\frac{h_n}{\sqrt{2}}}\right)^2\right)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-1}{2}\left(\frac{u-\mu}{\sigma}\right)^2\right)du$$

$$=\frac{1}{\left(\frac{h_n}{\sqrt{2}}\right)}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-1}{2}\left(\frac{x-u}{\left(\frac{h_n}{\sqrt{2}}\right)}\right)^2\right)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-1}{2}\left(\frac{u-\mu}{\sigma}\right)^2\right)du$$

$$=\frac{1}{\sqrt{2\pi\left(\frac{h_n^2}{2}+\sigma^2\right)}}\exp\left(\frac{-1}{2}\frac{(x-\mu)^2}{\frac{h_n^2}{2}+\sigma^2}\right)=\frac{\frac{h_n}{\sqrt{2}}}{2\pi\sqrt{\frac{h_n^2}{2}+\sigma^2}}\exp\left(\frac{-1}{2}\frac{(x-\mu)^2}{\frac{h_n^2}{2}+\sigma^2}\right)$$

$$\frac{1}{nh_n^2}E\left(\varphi^2\left(\frac{x-u}{h_n}\right)\right)=\frac{1}{nh_n 2\sqrt{2}\pi\sqrt{\frac{h_n^2}{2}+\sigma^2}}\exp\left(\frac{-1}{2}\frac{(x-\mu)^2}{\frac{h_n^2}{2}+\sigma^2}\right)$$

$$\lim_{h_n\to 0}\sqrt{\frac{h_n^2}{2}+\sigma^2}=\sigma \rightarrow \frac{1}{nh_n^2}E\left(\varphi^2\left(\frac{x-\mu}{h_n}\right)\right)=\frac{1}{nh_n}\cdot\frac{1}{2\sqrt{\pi}}\underbrace{\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)}_{P(x)}$$

(II) $$E^2\left(\varphi\left(\frac{x-u}{h_n}\right)\right)=\frac{h_n^2}{\sqrt{2\pi}\sqrt{h_n^2+\sigma^2}}\exp\left(\frac{-1}{2}\frac{(x-\mu)^2}{h_n^2+\sigma^2}\right)$$

$$\rightarrow \frac{1}{nh_n^2}E^2\left(\varphi\left(\frac{x-\mu}{h_n}\right)\right)\approx\frac{h_n}{nh_n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)\simeq 0$$

$$\rightarrow \mathrm{var}[P_n(x)]=\frac{1}{nh_n^2}E\left(\varphi^2\left(\frac{x-u}{h_n}\right)\right)\simeq\frac{1}{2nh_n\sqrt{\pi}}P(x)$$

2- باید بررسی کنیم که آیا خواص متریک $D(a,b)=\sqrt{\sum_{k=1}^{d}(a_k-b_k)^2}$ همچنان برقرار است یا نه.

$x'_k=\alpha_k x_k$

(A) خاصیت اول متریک این است که نامنفی باشد چون همچنان از جذر مجموع تعدادی مربعات جملات ساخته می‌شود پس همچنان نامنفی باقی می‌ماند.

ادامه سوال 2:

$$D'(a,b)=\sqrt{\sum_{k=1}^{d}\alpha_k^2(a_k-b_k)^2}$$

B - خاصیت بازتابی را چک می‌کنیم که اگر $D(a,b)$ برابر صفر باشد باید $a=b$ باشد.
از رابطه $D(a,b)$ واضح است که زمانی صفر می‌شود که $a=b$ باشد
همچنین زمانی که $a=b$ باشد $D(a,b)=0$ خواهد شد.

C - باید خاصیت تقارن را بررسی کنیم که $D(a,b)\overset{?}{=}D(b,a)$.

$$\sqrt{\sum_{k=1}^{d}\alpha_k^2(a_k-b_k)^2}=\sqrt{\sum_{k=1}^{d}\alpha_k^2(b_k-a_k)^2} \longleftarrow (a_k-b_k)^2=(b_k-a_k)^2$$

D - باید نامساوی مثلث را بررسی کنیم:

$$D(a,b)+D(b,c)\geq D(a,c)$$

$D'(a,b)=\|A(a-b)\|_2$ که A یک ماتریس قطری از ضرایب α_k است.

$$\|A(a-c)\|_2+\|A(c-b)\|_2\geq\|A(a-b)\|_2 \rightarrow \left\{ a-b=(a-c)+(c-b) \right.$$

به دنبال اثبات این هستیم.

$A(a-c)$ را x در نظر بگیریم و $A(c-b)$ را y و رابطه را بازنویسی می‌کنیم
هر دو طرف به توان ۲ می‌رسانیم

$$\|x\|_2+\|y\|_2\geq\|x+y\|_2 \rightarrow (\|x\|_2+\|y\|_2)^2 \geq \|x\|_2^2+\|y\|_2^2+2\|x\|_2\|y\|_2 \geq \|x\|_2^2+\|y\|_2^2+2\sum_{i=1}^{d}x_i y_i$$

$$\rightarrow \|x\|_2\|y\|_2\geq\sum_{i=1}^{d}x_i y_i$$ ← نیز این همان نامعادله کوشی شوارتز است

پس خاصیت نامساوی مثلث هم اثبات می‌شود.

چون این متریک همه عناصر متریک استاندارد را دارد پس می‌توان از آن برای knn استفاده کرد. چون به هر حال K عدد از فاصله ها کمتر از بقیه فواصل هستند و می‌توان حداکثر تعداد کلاس‌های نزدیک به یک داده جدید را به عنوان کلاس داده جدید در نظر گرفت.

ادامه سؤال 2 :

با توجه به اینکه ضریب α_k را در ابعاد ضرب می‌کنیم می‌توانیم با در نظر گرفتن $\alpha_k = 0$ برای بعضی ابعاد آن ها را حذف کنیم و روش پیش‌ساختاربندی و ویرایش را انجام دهیم. همچنین می‌توان عملیات حیاتی feature scaling را برای ابعاد مختلف انجام داد تا knn نتیجه منطقی‌تری به دست دهد.

---

(3) فرض می‌کنیم دسته اول مرکزش در 0 است و دسته دوم مرکزش روی 10 است

$$P(w_1) = P(w_2) = \frac{1}{2} \qquad P(w_1|x) = \begin{cases} 1 & \|x\| \leq 1 \\ 0 & o.w. \end{cases}$$

$$P(w_2|x) = \begin{cases} 1 & \|x - 10\| \leq 1 \\ 0 & o.w. \end{cases}$$

الف) $P_Q(e)$ متوسط احتمال خطا به ازای Q نقطه است:

$$P_Q(e) = P(x \in w_1 | w_2) + P(x \in w_2 | w_1) \xrightarrow{\text{تقارن شکل}}$$

$$P_Q(e) = 2P(x \in w_1 | w_2) = 2P(w_1)\, P\Big($$ برچسب w_1 برای کمتر از $\frac{(k-1)}{2}$ نقطه باشد و برچسب بقیه w_2 باشد $\Big)$

$$P_Q(e) = (2)\left(\frac{1}{2}\right) \sum_{j=0}^{\frac{k-1}{2}} P_0($$ j تا از Q نقطه w_1 برچسب خورده اند. $)$

$$P_Q(e) = \sum_{j=0}^{\frac{k-1}{2}} \binom{Q}{j} \frac{1}{2^j} \frac{1}{2^{Q-j}} = \frac{1}{2^Q} \sum_{j=0}^{(\frac{k-1}{2})} \binom{Q}{j}$$

ب) $P_Q(e,k)$ را به صورت متوسط احتمال خطا به ازای Q و k همسایگی تعریف می‌کنیم.

$$P_Q(e,1) = \frac{1}{2^Q} < P_Q(e,k) = \frac{1}{2^Q} \sum_{j=0}^{\frac{(k-1)}{2}} \binom{Q}{j} \quad (k>1)$$

c)

$$\lim_{Q\to\infty} P_Q(e) = 0$$

$$P_Q(e) = \frac{1}{2^Q} \sum_{j=0}^{\left(\frac{Q-1}{2}\right)} \binom{Q}{j} = P\left(B(n, \tfrac{1}{2}) \leq \tfrac{(k-1)}{2}\right)$$

$$= P\left(Y_1 + \cdots + Y_n \leq \tfrac{k-1}{2} \right]$$

که Y_i ها iid هستند . $P(Y_i=1) = P(Y_i=0) = \frac{1}{2}$ است. B یک توزیع دوجمله‌ای است.

$$P_Q(e) \leq P\left(Y_1 + \cdots + Y_n \leq \frac{\frac{a}{\sqrt{n}} - 1}{2} \right) \xrightarrow{n\to\infty} P(Y_1 + \cdots + Y_n \leq 0) = 0$$

$$\to \lim_{Q\to\infty} P_Q(e) = 0$$

سوال ۴

بخش اول

مفهوم مصالحه بایاس واریانس در knn و پارزن

در روش پارزن پارامتر hn نقش اساسی دارد. هرچقدر بزرگتر باشد مدل ساده تر خواهد بود و خطای بایاس بیشتر میشود و به نفع واریانس مصالحه انجام شده است. هرچقدر طول پنجره پارزن کوچکتر باشد مدل دقیقتر است و نویز داده هم ممکن است مدل شود و خطای وارياسن بیشتر خواهد شد.

در روش knn هایپرپارامتر ما k است. هرچقدر k که نشان دهنده تعداد داده موجود در هر حجم مورد نظر است میباشد کوچکتر باشد مدل اسپایکی تر خواهد شد و هرچقدر این تعداد همسایگی k بزرگتر بشود مدل نرم تر میشود. پس kکوچک خطای واریانس را زیاد میکند و مدل پیچیده تر میشود و k بزرگ خطای بایاس را زیاد میکند و مدل ساده تر میشود.

تفاوت روشهای پارامتری و غیرپارامتری

در روشهای پارامتری ما یک توزیع اولیه برای داده در نظر میگیریم و به روشهای مختلف به دنبال آن هستیم که پارامترهای این توزیع مفروض را تخمین بزنیم. اما در روشهای غیرپارامتری هیچ مدل اولیه ای برای داده فرض نمیشود و از خود توزیع داده ها برای تخمین نوع توزیع استفاده میشود. در واقع بر اساس همسایگی و هیستوگرام، توزیع داده ها را تخمین میزنیم. مزیت روش پارامتریک این است که حسبات بسیار سبک تر هستند و اگر از روش بیزی استفاده شود میتوان هربار نسبت به یک دیتاست جدید پارامتر توزیع را آپدیت کرد و بسیار خوش تعریف هستند. اما ایراد آن این است که همیشه توزیع داده ها مطابق توزیع مفروض ما نیست و میتواند تفاوت بسیار زیادی با ان داشته باشد. مزیت روشهای غیرپارامتری این است که مدل غیرواقعی برای توزیع فرض نمیکنند اما ایراد ان این است که همیشه به کل داده برای تخمین توزیع داده ها نیاز داریم و حجم محاسبات آن بسیار سنگین است.

مشکل روشهای بر اساس کرنل

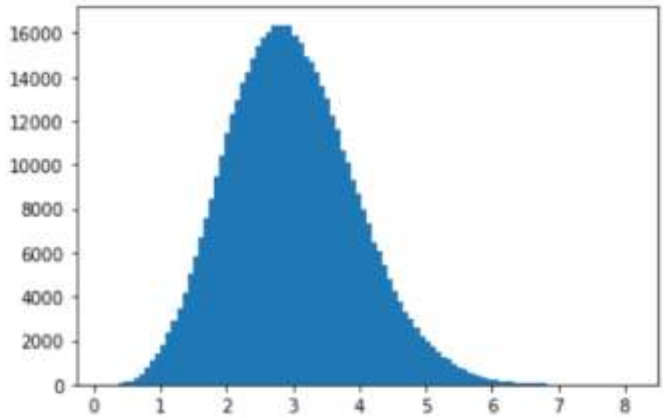
مشکل اصلی این است که اساسا تشخیص بهترین کرنل دشوار است و نمیتوان بهترین کرنل ممکن را در نظر گرفت. همچنین هنگامی که داده ها sparse باشند و در نواحی که داده های کمی وجود دارند این روش ها خوب عمل نمیکنند.

تفاوت مفهوم حجم در روش پارزن و knn

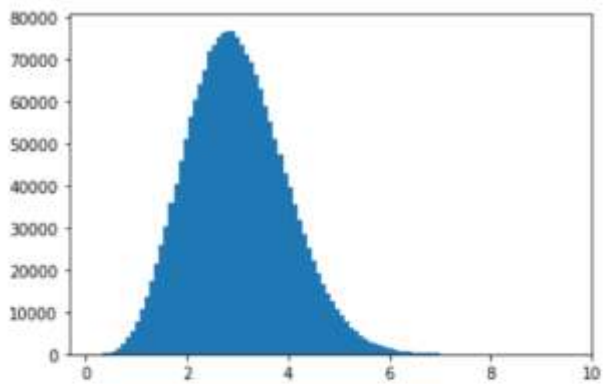
در روش پارزن حجم را برای یک تعداد مشخص داده ثابت فرض میکنیم و برای تخمین احتمال محاسبه میکنیم چه تعداد داده در هر حجم مفروض قرار گرفته است. این حجم با افزایش تعداد کل داده ها کوچکتر میشود. در روش knn ما تعداد داده که باید در یک حجم قرار بگیرد را برای یک تعداد مشخص ثابت فرض میکنیم و در اطراف داده جدید انقدر حجم را بزرگ میکنیم که تعدادk داده در آن حجم قرار بگیرد. سپس احتمال از فرمول k/nv محاسبه میشود. در این روش با افزایش تعداد کل داده k با سرعت کمتری افزایش پیدا میکند.

بخش دوم

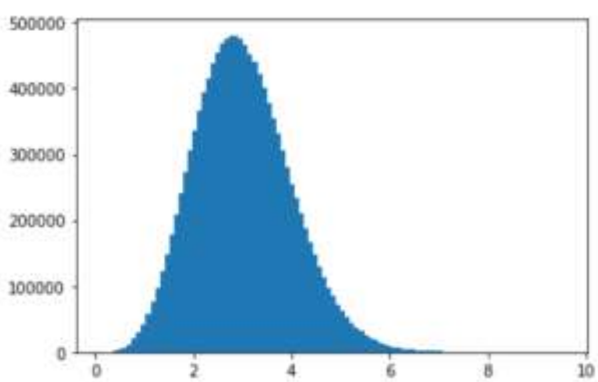
به سادگی یک ارایه ۱۰۰۰ در ۵ تولید میکنیم و سپس برای محاسبه فواصل از دو حلقه تو در تو استفاده میکنیم. نتایج برای داده های ۵ بعدی و مقدار داده ۱۰۰۰، ۲۰۰۰، ۵۰۰۰و ۱۰۰۰۰۰ در زیر رسم شده است.



N= 1000, d =5



N=2000 , d=5



N=5000,d=5

برای طولانی شدن کد هر بار مقدار n را تغییر دادم و دیگر برای هر n یک سلول در نظر نگرفتم.

ب)

سپس برای بررسی تاثیر افزایش ابعاد یک سری داده رندوم جدید به ابعاد ۱۰۰ در نظر گرفتم و این بار هم هر دفعه n را اپدیت کردم و هیستوگرامها ترسیم گردد.

متاسفانه به علت کندی بسیار در ران کردن کد و اینکه همانطور که در سایت ایلرن مشاهده میفرمایید بنده در آخرین ساعات تمرین را اپلود کردم و زمان کافی برای اجرای کد نبود. کدها هیچ ایرادی ندارند و صرفا به علت کمبود زمان نتوانستم نتیجه اجرای آنها را در گزارش بیاورم. لطفا نمره ای کسر نکنید. متشکرم.

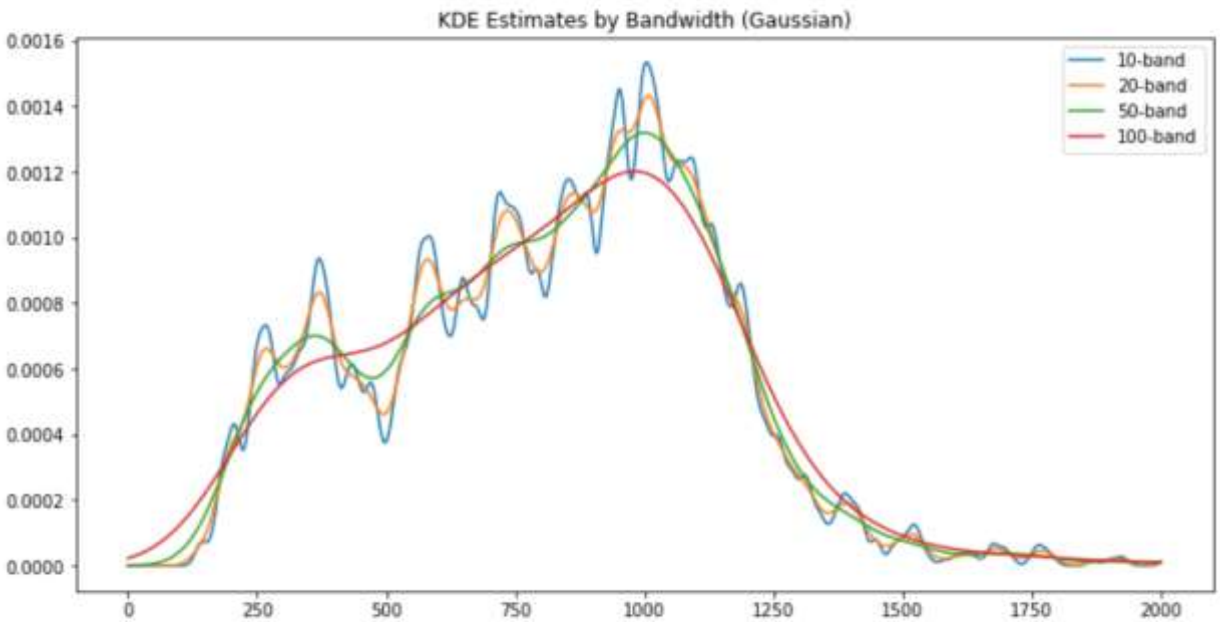
ج)

هرچقدر تعداد داده زیاد شد، هیستوگرام نرم تر میشود ولی به شدت نتیجه کندتر به دست می آید. در اثر افزایش ابعاد دچار نحصی ابعاد میشویم و تعداد داده مورد نیاز برای اموزش داده به صورت نمایی افزایش می یابد. خود افزایش ابعاد هم بسیار روند نتیجه گیری از کد را کند میکند. در روش knn که پیش پردازش داده ها در ان انجام ندشه باشد گاهی با تعداد خیلی زیادی داده سر و کار داریم و گاهی تعداد ابعاد این دادگان هم بسیار زیاد است. در این صورت الگوریتم knn بسیار زمان بر خواهد بود و هزینه محاسباتی زیادی خواهد داشت. در نتیجه اعمال روش های pre structuring برای knn ضروری به نظر میرسد.

سوال ۵

قسمت پیاده سازی بدون کتابخانه را نتوانستم انجام دهم.

ج) با استفاده از کتابخانه کرنل گاوسی را با پهنای باند ۱۰، ۲۰ ، ۵۰ و ۱۰۰ رسم کرده و در شکل زیر جواب نمایش داده شده است.



ب)

همانطور که از نمودار مشخص است با افزایش پهنای باند نمودار تخمین زده شده برای توزیع نرم تر میشود.

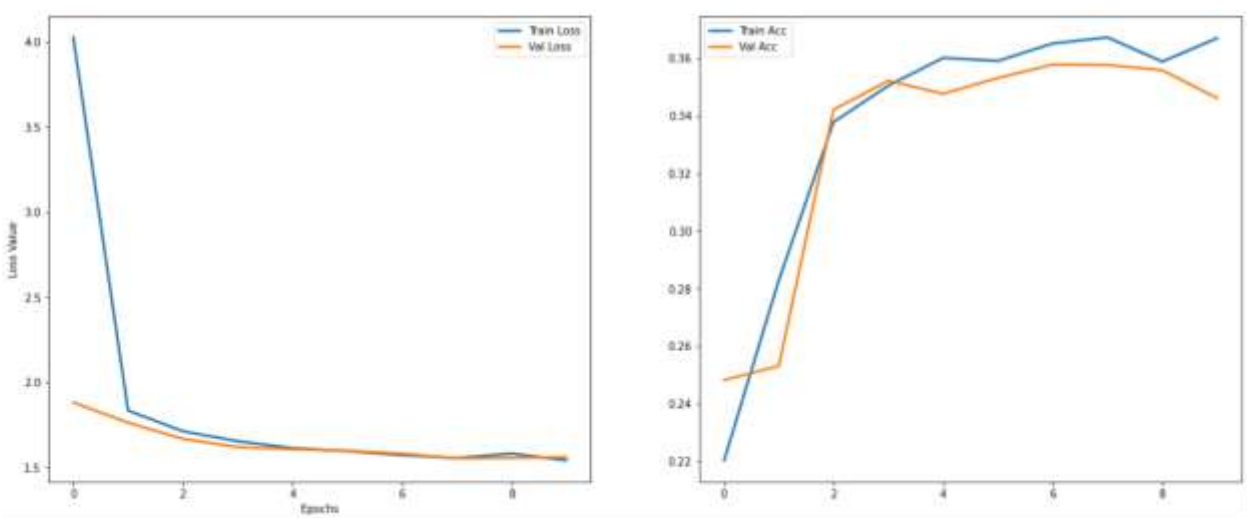
سوال ۶

ابتدا بخشی از داده را بعندان تست جدا میکنیم سپس در قسمت validation ، نسبت را ۰,۱ در نظر میگیریم.

مدل را sequential در نظر میگیریم و چون ورودی ها تصویر ها ۲۸ در ۲۸ هستند آنها را flatten میکنیم. برای learning rate یک بردار ۵ تای در نظر میگیریم و برای learning units هر لایه یک بردار ۳ تایی تعریف میکنیم. تابع activation را برای هر لایه relu در نظر گرفتم و برای لایه آخر که قرار است یکی از ۱۰ کلاس را اختصاص دهد تابع softmaxبهترین گزینه برای تابع فعال سازی بود. همچنین برای آموزش شبکه کل مدلها epoch را ۱۰ در نظر میگیریم و batch_size را ۱۶ تعیین میکنیم.

الف)

نمودار دقت و خطا برای داده train و validation را برای پارامترهای نرخ یادگیری 0/001 و یک لایه پنهان و تعداد ۵۰ یونیت و sover را sgd در نظر میگیریم. نتیجه در زیر نمایش داده شده است.



برای داده های تست loss = 1.56 و accuracy = 0.35 به دست آمد.

ب و ج)

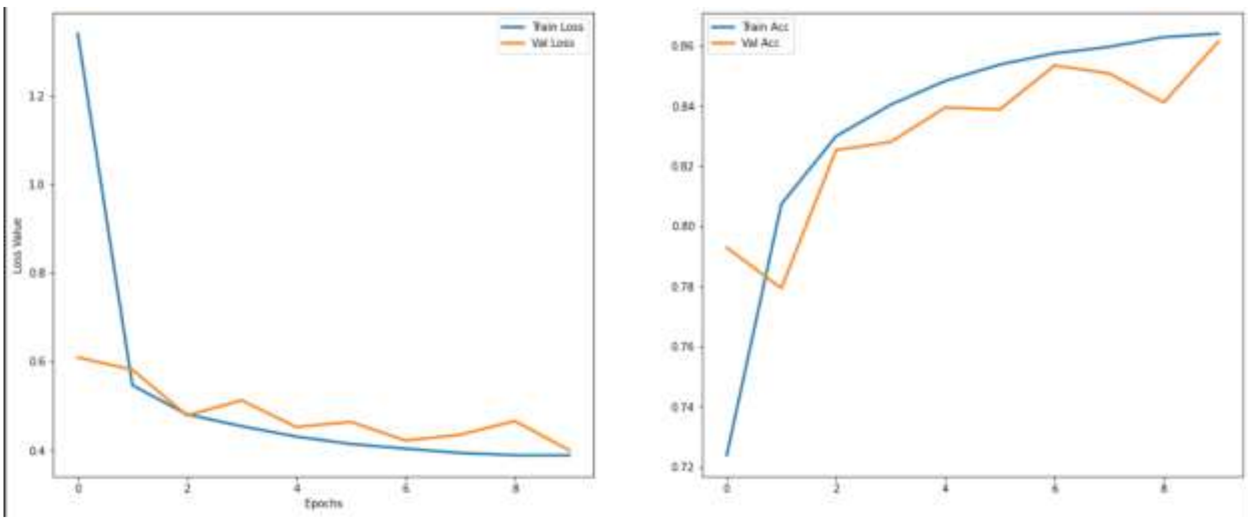
با بررسی learning rate بهترین مقدار نرخ یادگیری 0.001 به دست آمد. بقیه مقادیر تفاوت چندانی با هم نداشتند و الگوی مشخصی وجود نداشت چون در حالات مختلف هرکدام یک طور عمل میکردند.

هرچقدر تعداد لایه ها را افزایش دادم نتیجه بهتری گرفتم. برای بررسی این موضوع ابتدا لایه دوم و سوم را کامنت کردم و بهترین نتیجه با ۳ لایه پنهان به دست آمد.

در مورد اندازه یونیتهای لایه مخفی بهترین مقدار برابر ۱۵۰ شد و هرچقدر اندازه بزرگتر شد نتایج بهتری گرفتم.

بهترین نتیجه میان solver ها را با ADAM گرفتم. عملکرد sgd البته تفاوت زیادی با این نداشت اما اصلا نتایج خوبی با rpmprop حاصل نشد.

نتیجه استفاده از بهترین پارامترها در شکل زیر نمایش داده شده است.



برای داده های تست loss = 0.41 و accuracy = 0.85 به دست آمد.

سوال ۷

در این سوال ابتدا نقاط را در فضا تعریف کردم و یک بردار برای برچسبها تعریف کردم. سپس با line coordinates دو خط تعریف کردم با وزنهای که در انتها از آموزش شبکه به دست امدند.

تابعperceptron را تعریف کردم که در آن ورودی ها داده های اموزش، برچسبها،مقادیر اولیه رندوم وزن و بایاس و تعداد iteration هستند. خروجی آن وزنهای اصلاح شده، بایاس اصلاح شده و برچسبهای به دست آمده از شبکه جدید خواهد بود.

یک لایه آخر or برای ایجاد محیط مطلوب میان خطوط در نظر گرفته شد. سپس فرایند اموزش را مطابق انچه در فایل کد مربوطه است اجام دادیم. رابطه مربوط به پرسپترون را برای دو نورون لایه اول برای ورودی ها به ترتیب محاسبه میشود اگر برچسب درست اطلاق شد که میرود ورودی بعدی. در غیر اینصورت اگر برچسب

۱ باشد، وزن و بایاس متناظر با کمترین zin را اپدیت میکند. و اگر برچسب -۱ باشد zin که مثبت را با تابع posefinderپیدا کرده و وزن مربووطه را اصلاح میکند. در انتها برچسب نهایی به داده اطلاق میشود. یعنی برچسبی که غلط بوده را اصلاح میکنیم.

معادله خطوط اولیه جدا کننده پس از اموزش شبکه بعد از اصلاح ماتریس وزنها و بایاس به دست می اید و نتیجه در شکل زیر نمایش داده شده است.