Subject Date: Sa Su Mo Tu We Th Fr

1)

الف) برای کلاس k ام اثبات می‌کنیم نتیجه قابل تعمیم به همه کلاس‌ها است.

$$P(Error) = \sum_j P(y=k, x\in R_j), j\neq k + \sum_i P(y=i, x\in R_k), i\neq k + P(y=k, x\in R_k)$$

$$= P(y=k) + \underbrace{\left(\sum_i P(y=i, x\in R_k), i\neq k - P(y=k, x\in R_k)\right)}_{A}$$

$$A = \int_{R_k} \Big(\underbrace{\sum_i P(x, y=i)}_{P(x)P(y=i|x)} - \underbrace{P(x, y=k)}_{P(x)P(y=k|x)} \Big) dx = \int_{R_k} P(x) \Big(\underbrace{\sum_i P(y=i|x)}_{II} - \underbrace{P(y=k|x)}_{I} \Big) dx$$

← روی ناحیه k ام (I) بزرگ‌تر از (II) است یعنی $P(y=k|x)$ در ناحیه R_k بزرگ‌تر از باقی کلاس‌ها در آن ناحیه است.

با این فرض A منفی است و کاهنده خطا است. این قابل تعمیم به تمام کلاس‌ها است لذا در مجموع خطا کمینه می‌شود.

ب)

نیز کلاسی را انتخاب می‌کند که $P(y=i|x)$ آن از همه بیشتر باشد. حال اگر M کلاس داشته باشیم فضا به M قسمت افراز می‌شود لذا $\sum_{i=1}^{M} P(y=i|x) = 1$

حال چون $\sum_{i=1}^{M} P(y=i|x) = 1$ پس باید حداقل یکی از آن‌ها زیر سیگما بزرگ‌تر مساوی $\frac{1}{M}$ باشد.

حال اگر فرض کنیم کلاس k ام انتخاب شده:

$$P(y=k|x) = \max_i P(y=i|x) \geq \frac{1}{M} \quad (I)$$

$$\Rightarrow P(Error) = 1 - P(y=k|x) \overset{(I)}{\Rightarrow} \leq 1 - \frac{1}{M} = \frac{M-1}{M}$$ ✓

ج) دو روش پیشنهاد می‌کنیم:

روش اول: برای هر کلاس به طور 1vs all نگاه کنیم و یک ROC بکشیم.

روش دوم: برای هر th تمام کلاس‌ها را دو به دو در نظر می‌گیریم، در نهایت روی همه اعداد به دست آمده میانگین می‌گیریم و خروجی برای آن th به دست می‌آید. این کار را برای همه th ها تکرار می‌کنیم.

د) فرضی که در naive bayes می‌کنیم استقلال متون هاست حال اگر مجموعه داده‌ای وجود داشته باشد که بردارها ویژگی مختلف در آن مستقل باشند naive bayes طبقه‌بند بهینه است.

CamScanner

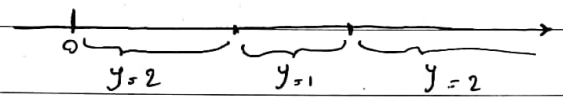
Subject: Date: Sa Su Mo Tu We Th Fr

2)

با مساوی قرار دادن توزیع ها داده شده مرز تصمیم به دست می آید.

$$\frac{x}{a^2}\exp\left(-\frac{x^2}{2a^2}\right) = \theta x \exp(-\theta x) \xrightarrow{\ln} -2\ln a - \frac{x^2}{2a^2} = \ln\theta - \theta x$$

$$\Rightarrow x^2 - 2a^2\theta x + 2a^2(\ln\theta a^2) = 0 \Rightarrow x = \theta a^2 \pm a\sqrt{a^2\theta^2 - 2\ln\theta a^2}$$



3)

الف)

$$risk = \int_{R_1} \lambda_{12} P(\omega_2) P(x|\omega_2) dx + \int_{R_2} \lambda_{21} P(\omega_1) P(x|\omega_1) dx$$

$$= \lambda_{12}\int_{R_1} P(x|\omega_2)dx + \lambda_{21}\int_{R_2} P(x|\omega_1)dx$$ با فرض $P(\omega_2) = P(\omega_1)$.

$$\xrightarrow{\lambda_2 = \lambda_{21} = 1} \int_{R_2} P(x|\omega_1)dx = \int_{R_1} P(x|\omega_2)dx \checkmark$$

ب) بدون فرض $P(\omega_1) = P(\omega_2)$ قسمت الف صحیح نخواهد بود لذا همواره پاسخ یکتا نخواهد بود و اگر این شرط برقرار نباشد نتیجه قسمت الف درست نیست.

4)

برای حل این سوال با استفاده از ln مسئله به سادگی حل می‌شود.

$$\overset{A}{\ln(P(\mu|x))} = \sum_{k=1}^{N} \ln(P(x_k|\mu)) + \ln(P(\mu)) = -\sum_{k=1}^{N}\frac{(x_k-\mu)^2}{2a^2} + \ln\mu - \frac{\mu^2}{2a_\mu^2} - \ln a_\mu^2$$

$$\frac{\partial A(\mu)}{\partial\mu} = -\sum_k \frac{\mu - x_k}{a^2} + \frac{1}{\mu} - \frac{\mu}{a_\mu^2} = 0 \Rightarrow \dots \Rightarrow \left(\frac{N}{a^2} + \frac{1}{a_\mu^2}\right)\mu^2 - Z\mu - 1 = 0$$

$$\Rightarrow \hat{\mu}_{MAP} = \frac{Z}{2R}\left(1 + \sqrt{1 + \frac{4R}{Z^2}}\right) \checkmark$$

5)

الف)

$$L(\theta|D) = P(D|\theta) = \prod_{i=1}^{n} P(y_i|\theta), \quad \hat{\theta}_{ml} = \arg\max_{\theta} L(\theta|D) = \arg\max_{\theta} \sum_{i=1}^{n} \ln(P(y_i|\theta))$$

$$\Rightarrow \nabla_\theta \ln(P(y_i|\theta)) = 0 \Rightarrow \sum_{i=1}^{n} -\frac{1}{\theta} + \frac{y_i^2}{\theta^2} = 0 \Rightarrow \hat{\theta}_{ml} = \frac{1}{n}\sum_{i=1}^{n} y_i^2$$

ب) وقتی بی نهایت مشاهده داشته باشیم MAP به MLE میل می‌کند.

CamScanner