中国作家协会

Past mid term:

#1. Show the mode (maximum) of the multivariate Gaussian distribution. $N(x|\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{1}{2\sigma^2}(x-\mu)^2)$ is given by μ.

Solution: fixed σ^2, to maximize N's to maximize $\exp(-\frac{1}{2\sigma^2}(x-\mu)^2) \Rightarrow$ maximizing $-\frac{1}{2\sigma^2}(x-\mu)^2 \Rightarrow$ minimize $(x-\mu)^2 \Rightarrow x=\mu$.

#2. Sps dataset of observations $\vec{x}=(x_1,x_2,\cdots,x_N)^T$, N observations of scalar binary ~~ob~~ variable x ($x\in\{0,1\}$). Assume obs. iid drawn from Bernoulli (μ)

(a). log-likelihood for N obser. $L=\prod_{i=1}^N \mu^{x_i}(1-\mu)^{1-x_i} \Rightarrow \log L=\sum_{i=1}^N[x_i\log\mu+(1-x_i)\log(1-\mu)]$

(b). find μ_{ML}. Let $M=$ # of 1's in observations, $M=\sum_{i=1}^N x_i$. $\operatorname{Log}L=M\log\mu+(N-M)\log(1-\mu)$ $\frac{\partial \log L}{\partial\mu}=\frac{M}{\mu}-\frac{N-M}{1-\mu}=0 \Rightarrow \mu_{ML}=\frac{M}{N}$.

check 2nd derivative <0 at $\mu_{ML} \Rightarrow \checkmark$.

(c). Show $E[\mu_{ML}]=\mu$

$E[\mu_{ML}]=E(\sum x_i/N)=\frac{1}{N}\sum_{i=1}^N E(x_i)=\frac{1}{N}N\mu=\mu$ ✓

#3. dataset of observ. $\vec{x}=(x_1,\cdots,x_N)^T$ drawn from $N(\mu,\sigma^2)$, sps $\sigma^2_{ML}=\frac{1}{N}\sum_{n=1}^N(x_n-\mu_{ML})^2$, but replace μ_{ML} with μ. show $E[\sigma^2_{ML}]=\sigma^2$

$E(\sigma^2_{ML})=E(\frac{1}{N}\sum_{i=1}^N(x_i-\mu)^2)=\frac{1}{N}[\sum_{i=1}^N[E(x_i-\mu)^2]]$

$E[(x_i-\mu)^2]=E[(x_i-E(x_n))^2]$ by the def of var, $E[(x_n-\mu)^2]$ is the true variance σ^2 for x. $E[(x_i-\mu)^2]=\sigma^2\ \forall i$. Thus $E(\sigma^2_{ML})$ $=\frac{1}{N}\sum_{i=1}^N\sigma^2=\sigma^2$.

#4. reg model $y(\vec{x})$ with square loss func, expected loss: $E[L]=\iint(y(\vec{x})-t)^2p(\vec{x},t)\,d\vec{x}\,dt$. Show the function $y(\vec{x})$ for which this expected loss is $\downarrow$ by the conditional expectation $E[t|\vec{x}]=\int tp(t|\vec{x})dt$.

Prof: $E[L]=\iint y(x)^2p(x,t)\,dx\,dt+\iint t^2p(x,t)dx\,dt-2\iint y(x)\,t\,p(x,t)dx\,dt=\int y(x)^2p(x)dx+\int t^2p(t)dt-2\int y(x)dx\int tp(t|x)p(x)dt$

$=\int y(x)^2p(x)dx+E[t^2]-2\int y(x)p(x)E(t|x)dx=E_t(t^2)+E_x[y(x)^2-2y(x)E(t|x)]$

To minimize $E[L]$ is to minimize $E_x(\quad)$, $\forall x$, $y(x)^2-2E(t|x)y(x)$ is $\downarrow$ at $y(x)=E(t|x) \Rightarrow$

#5. linear basis func. reg model. $X=\{x_1,\cdots,x_N\}$, target value $t=\{t_1,\cdots,t_N\}$. Assume iid obs., likelihood func.

$p(t|X,w,\beta)=\prod_{n=1}^N N(t_n|w^T\phi(x_n),\sigma^2)$. Consider zero mean Gaussian prior governed by single precision parameter

α: $p(w|\alpha)=N(w|0,\alpha^{-1}I)$ $\angle$ marginal likelihood?

(a). $p(t|X,\alpha,\beta)=\int p(t|X,w,\beta)p(w|\alpha)dw=\int\prod_{n=1}^N\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{1}{2\sigma^2}(x_n-w^T\phi(x_n))^2)\cdot\frac{1}{\sqrt{2\pi\sigma^2}}\exp(\cdots\cdots)$

(b). predictive distribution of t for x^*.

(c). Show maximizing the 'log' of posterior distribution wrt w is $\Leftrightarrow$ minimizing the sum-of-squares error function, with addition to a quadratic regularization term with $\lambda=\alpha\sigma^2$

posterior: $p(w|D)$ since $p(w|D)\propto p(w)\cdot p(D|w)$

$p(w|D)=C\cdot\exp(-\frac{\beta}{2}\sum_{n=1}^N(t_n-w^T\phi(x_n))^2]\cdot\exp(-\frac{\alpha}{2}w^Tw)$, c constant

$$\underset{\max}{\log p(w|D)}=\underbrace{-\frac{\beta}{2}\sum_{n=1}^N(t_n-w^T\phi(x_n))^2-\frac{\alpha}{2}w^Tw}_{\min}+\text{constant}$$

$$=\beta[\underbrace{\tfrac{1}{2}\sum_{n=1}^N(t_n-w^T\phi(x_n))^2}_{\text{sum of square error}}+\underbrace{\tfrac{1}{2}\tfrac{\alpha}{\beta}w^Tw}_{\text{quadratic term with }\lambda=\frac{\alpha}{\beta}=\alpha\sigma^2}]\quad(\beta=(\sigma^2)^{-1}>\text{precision}$$

#6. Bayesian logistic reg for binary classification, parametrized by w. $\{x_n,t_n\}$, $t_n\in\{0,1\}$, $n=1,\cdots,N$. zero mean Gaussian prior over w governed by a single precison ~~pa~~ α: $p(w|\alpha)=N(w|0,\alpha^{-1}I)$

(a). posterior distribution over w for Bayesian logistic regression:

$p(w|x,t)\propto p(t|x,w)\,p(w|\alpha)$. $p(w|\alpha)=N(w|0,\alpha^{-1}I)$

$\log p(w|x,t)=\log p(t|x,w)+\log p(w|\alpha)+\text{constant}=-\frac{1}{2}w^T\alpha Iw+\sum_{n=1}^N[t_n\ln y_n+(1-t_n)\ln(1-y_n)]+\text{cons.}$

$\Rightarrow p(w|x,t)\propto\exp\{-\frac{1}{2}\alpha w^Tw\}\cdot\{\prod_{i=1}^N y_n^{t_n}(1-y_n)^{(1-t_n)}$ where $y_n=\sigma(w^Tx_n)=\frac{1}{1+\exp(-w^Tx_n)}$

(b). Show $\uparrow$ ~~p~~log of posterior distribution $\Leftrightarrow\downarrow$ cross-entropy error function with addition quadratic regularization term.

Similar to 5(b): $\uparrow\log p(w|x,t)\Leftrightarrow\downarrow-\sum[\quad]+\frac{\alpha}{2}w^Tw=\sum_{n=1}^N E_n+\frac{\alpha}{2}w^Tw$. $E_n=\begin{cases}-\ln y_n & \text{if } t_n=1\\ -\ln(1-y_n) & \text{if } t_n=0\end{cases}$. $\sum E_n$ is cross-entropy error in total.

$\downarrow$ & $\lambda=\alpha$. as regularization coefficient.

(c). Compute deriv of ~~E~~ cross-entropy wrt w.

~~$E[L]=-\sum_{n=1}^N\{t_n\ln y_n+(1-t_n)\ln(1-y_n)\}+\frac{\alpha}{2}w^Tw=\sum_{n=1}^N E_n+\frac{\alpha}{2}w^Tw$~~ $\nabla_w[w^Tw]=2w$

$\frac{dE_n}{dy_n}=\frac{y_n-t_n}{y_n(1-y_n)}$ $\frac{dy_n}{dw}=y_n(1-y_n)x_n$ since $\frac{dy_n}{dw}=\frac{d\sigma(w^Tx_n)}{dw}=y_n(1-y_n)x_n \Rightarrow \frac{dE_n}{dw}=(y_n-t_n)x_n \Rightarrow \nabla_wE[L]=\sum_{n=1}^N(y_n-t_n)x_n+\alpha w$

Selected problems.

中国作家协会

.10. x, z independent. Show mean & variance of ~~their sum satisfies $E(x+z)=E(x)+E(z)$, $V(x+z)=V(x)+V(z)$~~.

$$\left.\begin{aligned}&(x+z)=\iint(x+z)p(x)p(z)\,dx\,dz=\int x p(x)dx+\int z p(z)dz=E(x)+E(z).\\&(x+z)=\iint(x+z-E(x+z))^2\,\cancel{dxdz}\,p(x)p(z)\,dx\,dz=\int(x-E(x))^2p(x)dx+\int(z-E(z))^2p(z)dz=var(x)+Var(z)\end{aligned}\right\}$$

continuous
if discrete use

.8. $p(x,y)$. prove $E(x)=E_y[E_x[x|y]]$, $var(x)=E_y[var_x[x|y]]+var_y[E_x[x|y]]$

product rule: $E_y[E_x[x|y]]=\int\{\int xp(x|y)dx\}p(y)dy=\iint xp(x,y)\,dx\,dy=\int xp(x)dx=E_x[x]$

$E_y[\cancel{E_x}\,Var_x[x|y]]+Var(E_x[x|y])=E_y[E_x[x^2|y]-E_x[x|y]^2]+E_yE_x[x|y]^2]-E_y[E_x[x|y]]^2=\cdots$ ✓