作业3

2017年4月13日　20:34

1. $\sigma^2(1-\frac{d+1}{N}) = 0.1^2(1-\frac{8+1}{N}) > 0.008 \Rightarrow N > 45$

2. (a) 正确　$y^T H y = y^T \hat{y}$　y与y的投影$\hat{y}$夹角小于等于$90°$

故　$y^T H y \geq 0$　H是半正定的

(b) 错误，如果H可逆，由于$H^2=H$，则 $H^2H^{-1}=HH^{-1}$

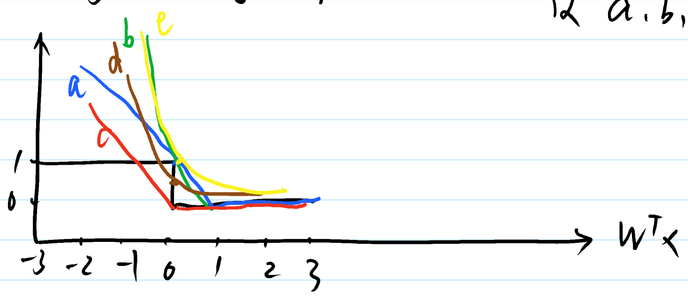
$\Rightarrow H=I$

(c) 错误：$Hy=\lambda y=\hat{y}$，$\hat{y}$是y的投影，故只有可能是 $\lambda=1$

(d) 正确

(e) 正确　多次投影相当于1次投影

3. 以 $y=1$ 为例绘图　　仅 a.b，e 可作为上限



4. b，d，e 处处可微

5. $err(w)=\max(0,-yw^Tx)$　不失一般性假设 $y=1$

① $w^Tx<0$　$err(w)=-yw^Tx$　$\nabla err=-yx$　$y\neq sign(w^Tx)$

② $w^Tx>0$　$err(w)=0$　$\nabla err=0$　$y=sign(w^Tx)$

代入SGD的公式

$$w_{t+1}\leftarrow w_t-\eta\nabla err$$

令 $\eta=1$，则 $w_{t+1}\leftarrow w_t-\nabla err = w_t+[\![y\neq sign(w^Tx)]\!](yx)$

易见为PLA算法

6. $\nabla E(u,v)=(e^u+ve^{uv}+2u-2v-3,\ 2e^{2v}+ue^{uv}-2u+4v-2)$

$\nabla E(u,v)_{u=0,v=0}=(-2,0)$

7. $(u_{t+1},v_{t+1})=(u_t,v_t)-\eta\nabla E(u,v)$

使用6的结果迭代计算，易得 $E(u_5,v_5)=2.825$　参见代码

8. $\frac{\partial^2E}{\partial u^2}=e^u+v^2e^{uv}+2$　$\frac{\partial^2E}{\partial v^2}=4e^{2v}+u^2e^{uv}+4$　$\frac{\partial^2E}{\partial u\partial v}=e^{uv}+uve^{uv}-2$

8, $\frac{\partial^2 E}{\partial u^2} = e^u + v^2 e^{uv} + 2 \quad \frac{\partial^2 E}{\partial v^2} = 4e^{2v} + u^2 e^{uv} + 4 \quad \frac{\partial^2 E}{\partial u \partial v} = e^{uv} + uve^{uv} - 2$

$b_{uu} = \frac{1}{2}\frac{\partial^2 E}{\partial u^2}(0.0) = 1.5 \quad b_{vv} = \frac{1}{2}\frac{\partial^2 E}{\partial^2 v}(0.0) = 4$

$b_{uv} = \frac{\partial^2 E}{\partial u \partial v}(0.0) = -1 \quad b_u = \frac{\partial E}{\partial u}(0.0) = -2 \quad b_v = \frac{\partial E}{\partial v}(0.0) = 0$

$b(u.v) = E(0.0) = 3$

9. 方向为 $-(\nabla^2 E(u.v))^{-1} \nabla E(u.v)$

$\Delta u . \Delta v \to 0$ 时

$b_{uu}(\Delta u)^2 + b_{vv}(\Delta v)^2 + b_{uv}\Delta u \Delta v = -(b_u \Delta u + b_v \Delta v)$

即 $[\Delta u, \Delta v] \begin{bmatrix} \frac{\partial^2 E}{\partial v^2} & \frac{\partial^2 E}{\partial u \partial v} \\ \frac{\partial^2 E}{\partial u \partial v} & \frac{\partial^2 E}{\partial v^2} \end{bmatrix} [\Delta u . \Delta v]^T = -[\frac{\partial E}{\partial u}, \frac{\partial E}{\partial v}) (\Delta u . \Delta v)^T$

$\Rightarrow [\Delta u . \Delta v] = -(\nabla^2 E(u.v))^{-1} \nabla E(u.v)$

10, 迭代计算得 $E(u_5\ v_5) = 2.361$ 参见代码

11. 考虑变换 $\Phi_2(x) = (1, x_1, x_2, x_1^2, x_1x_2, x_2^2)$

$X \xrightarrow{\Phi_2(x)} Z$，对任意 y $\quad Zw = sign(y)$ 取 $Z\tilde{w} = y$ 即可

如 Z 可逆，则可以得到 $\tilde{w} = Z^{-1}y$ 得到 shatter 的参数

它是二次曲线，直线，常数的组合

12, N 个样本在 Z 空间中转换为 N 阶单位阵

取 $\tilde{w} = y$ 即可 shatter，N 可以任意大，故 X 空间中 VC 维为 ∞

13～15 参见代码 线性分割时 $E_{in} = 0.51$

二次分割时 $w_3 = 0.001 \quad E_{out} = 0.13$

16. $\max\limits_{w}$ likelihood$(w) \propto \prod\limits_{n=1}^{N} h_y(x_n)$

$$\max_{w} \ln \prod_{n=1}^{N} \frac{\exp(w_{y_n}^T x_n)}{\sum_{i=1}^{k} \exp(w_i^T x_n)}$$

$$\min_{w} \frac{1}{N} \sum_{n=1}^{N} -\ln \frac{\exp(w_{y_n}^T x_n)}{\sum_{i=1}^{k} \exp(w_i^T x_n)}$$

$$E_{in} = \frac{1}{N} \sum_{n=1}^{N} \ln\left(\sum_{i=1}^{k} \exp(w_i^T x_n)\right) - w_{y_n}^T x_n)$$

17. $\frac{\partial E_{in}}{\partial w_i} = \frac{1}{N} \sum_{n=1}^{N} \left(\frac{\exp(w_i^T x_n)}{\sum^{k} \exp(w_i^T x_n)} x_n - [[y_n = i]] x_n \right)$

17. $$\frac{\partial E_{in}}{\partial w_i} = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{\exp(w_i^T x_n)}{\sum_{i=1}^{k}\exp(w_i^T x_n)} x_n - [[y_n=i]]x_n\right)$$

$$= \frac{1}{N}\sum_{n=1}^{N}\left(\left(h_i(x_n) - [[y_n=i]]\right)x_n\right)$$

18 ~ 20 编程题

18. 0.47

19. 0.22

20. 0.47