TDA 231

EXAM 2016

SOLUTION SKETCHES

Problem 1

(a) $P[(x_i, y_i)\ i=1,\dots,N \mid m, \sigma^2]$

$$= \prod_{i=1}^{N} P[(x_i, y_i) \mid m, \sigma^2]$$

$$= \prod_{i=1}^{N} \mathcal{N}(y_i - m x_i \mid 0, \sigma^2)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left[\frac{\sum_{i=1}^{N} (y_i - m x_i)^2}{2\sigma^2}\right]$$

(b) Taking logs and differentiating

$$m_{MLE} = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2}$$

(c) Conjugate prior:

$$\mathcal{N}(m \mid m_0, \sigma_0^2)$$

(d) Multiplying prior & likelihood and taking only exp term:

$$\exp\left[\frac{\sum_{i=1}^{N}(y_i - m x_i)^2}{2\sigma^2} + \frac{(m-m_0)^2}{2\sigma_0^2}\right]$$

which by conjugacy should equal

$$\exp\left[\frac{(m-\hat{m})^2}{2\hat{\sigma}^2}\right]$$

Compare coefficients of m^2 and m:

$$\frac{\sum_i x_i^2}{2\sigma^2} + \frac{1}{2\sigma_0^2} = \frac{1}{2\hat{\sigma}^2}$$

$$\frac{\sum_i x_i y_i}{\sigma^2} + \frac{m_0}{\sigma_0^2} = \frac{\hat{m}}{\hat{\sigma}^2}$$

These two equations give $(\hat{m}, \hat{\sigma})$ which gives us the posterior dist:

$$\mathcal{N}(m \mid \hat{m}, \hat{\sigma}).$$

Problem 2

(a) $P[Y = c, X_1 = x_1, X_2 = x_2]$

$$= P[Y=c]\, P[X_1 = x_1 \mid Y = c]\, P[X_2 = x_2 \mid Y = c]$$

$$= \pi_c\, \theta_c^{x_1} (1-\theta_c)^{1-x_1}\, \mathcal{N}(x_2 \mid \mu_c, \sigma_c^2)$$

(b) $P[Y = c \mid X_1 = 0, X_2 = 0]$

$$= \frac{P[X_1 = 0 \mid Y = c]\, P[X_2 = 0 \mid Y = c]\, P[Y = c]}{P[X_1 = 0, X_2 = 0]}$$

$$= \frac{P[X_1 = 0 \mid Y = c]\, P[X_2 = 0 \mid Y = c]\, P[Y = c]}{\sum_{c'} P[X_1 = 0 \mid Y = c']\, P[X_2 = 0 \mid Y = c']\, P[Y = c']}$$

Now substitute into this for $c = 1, 2, 3$.

(Note: Strictly speaking $P(X_2 = 0 \mid Y = c) = 0$ since X_2 is continuous. However, this is OK since one can take a small interval around 0 and use the pdf instead.)

(c) & (d) are similar.

Problem 3

(a) PGMs & plate notation not covered this year.

(b) $P[(x_i, y_i), i = 1, .., N]$

$$= \prod_{i=1}^{N} P[(x_i, y_i)]$$

$$= \prod_{i=1}^{N} P[\mathcal{L} = \ell] \, P[x_i \mid \mathcal{L} = \ell] \, P[y_i \mid x_i, \mathcal{L} = \ell]$$

$$= \prod_{i=1}^{N} \frac{1}{L} \mathcal{N}(x_i \mid \mu_\ell, 100) \, \mathcal{N}(y_i - m_\ell x_i \mid 0, \sigma^2)$$

(c) Suppose σ^2 is known constant. Then the parameter space is $((\mu_\ell, m_\ell), \ell = 1, .., L)$.

(d) A Markov chain on this state space can be defined by transitions of the form: for each $\ell \in 1, .., L$

$$(\mu_\ell, m_\ell) \longrightarrow (\mu_\ell', m_\ell)$$

$$(\mu_\ell, m_\ell) \longrightarrow (\mu_\ell, m_\ell')$$

Denote the probability of the likelihood function above as P.

The proposals

$$(\mu_e, m_e) \rightarrow (\mu_e', m_e)$$

can be made by adding a random step:

$$\mu_e' = \mu_e + \mathcal{N}(0, \sigma'^2) \quad \text{etc}$$

and accepted with probability

$$\min\left(1, \frac{P'}{P}\right)$$

Note that in the ratio P'/P, everything cancels except

$$\mathcal{N}(x_i \mid \mu_e', 100) / \mathcal{N}(x_i \mid \mu_e, 100)$$

and in the case of m_e',

$$\mathcal{N}(y_i - m_e x_i \mid 0, \sigma^2) / \mathcal{N}(y_i - m_e' x_i \mid 0, \sigma^2)$$

Problem 4

(a) In the E-step assign point (x_i, y_i) to the line $y = m_\ell x$ with probability

$$\frac{e^{\Delta_\ell}}{\sum_{\ell'} e^{\Delta_{\ell'}}}$$

(b) In the M-step, assuming points have been assigned to lines, compute m_ℓ as the least-squares estimate of the points assigned to that line (or MLE estimate).

(c) Initialize with m_ℓ assigned at random. The output will depend on the initialization.

(d) MCMC will give more precise posterior distribution information indep of initialization but is computationally heavier

Problem 5

$$\bar{h} := \sigma(W^{(1)}x + b^{(1)})$$

$$\hat{y} := \text{softmax}(W^{(2)}h + b^{(2)})$$

Introduce

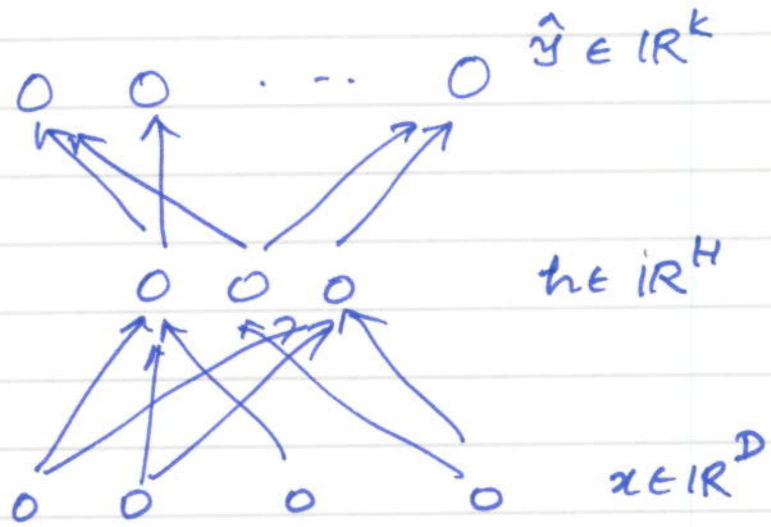
$$z^{(1)} := W^{(1)}x + b^{(1)}$$
$$z^{(2)} := W^{(2)}h + b^{(2)}$$

So

$$\hat{y} = \text{softmax}(z^{(2)})$$

$$\bar{h} = \sigma(z^{(1)})$$

$$C(y, \hat{y}) := -\sum_i y_i \log \hat{y}_i$$



(a) $$W^{(1)} \in \mathbb{R}^{H \times D}$$
$$b^{(1)} \in \mathbb{R}^{H}$$
$$W^{(2)} \in \mathbb{R}^{K \times H}$$
$$b^{(2)} \in \mathbb{R}^{K}$$

$HD + KH + H + K$

$= (D+1)H + (H+1)K$

parameters

(b) $$\frac{\partial C}{\partial z^{(2)}} = y - \hat{y}$$ [From your assignment]

$$\frac{\partial C}{\partial h_i} = \sum_k \frac{\partial C}{\partial z_k^{(2)}} \frac{d z_k^{(2)}}{d h_i}$$

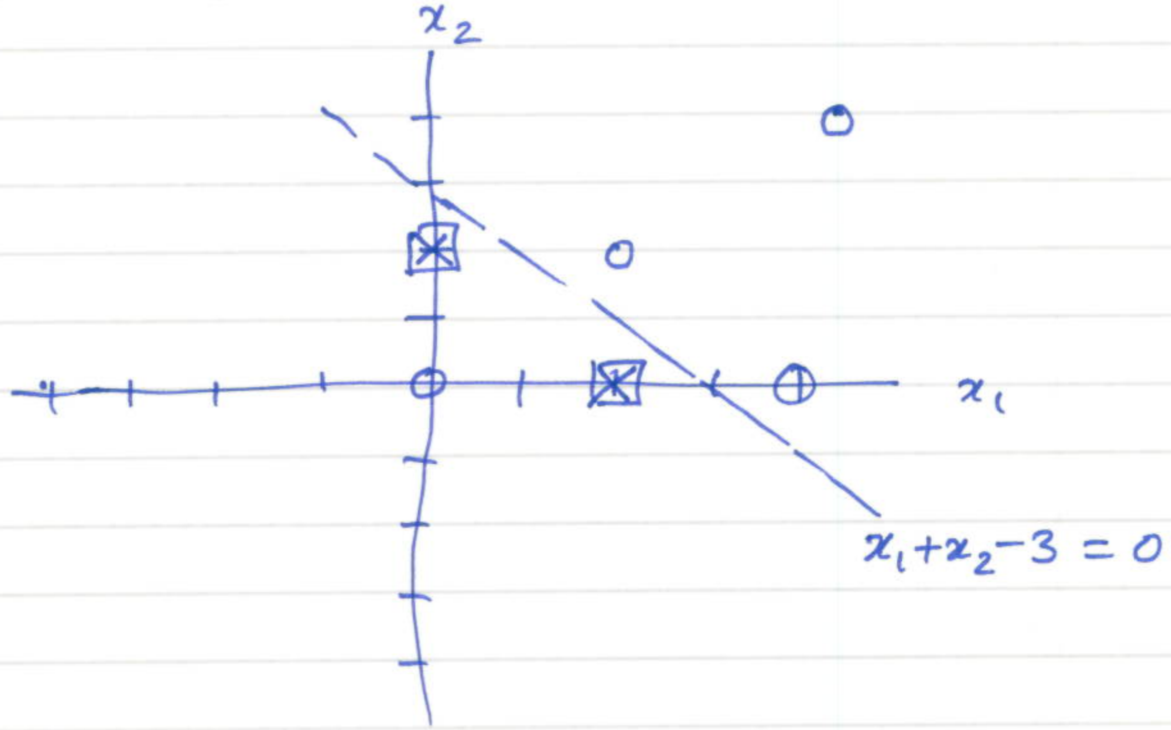
$$= \sum_k (y_k - \hat{y}_k) w_{i,k}^{(2)}$$

$$\frac{\partial C}{\partial z^{(1)}} = \frac{\partial C}{\partial h_i} \frac{d h_i}{d z_i^{(1)}} = \frac{\partial C}{\partial h_i} h_i(1-h_i)$$

$$\frac{\partial C}{\partial w_{ij}^{(1)}} = \frac{\partial C}{\partial z_i^{(1)}} \frac{d z_i^{(1)}}{d w_{ij}^{(1)}} = \frac{\partial C}{\partial z^{(1)}} \cdot x_j$$

$$= \sum_k (y_k - \hat{y}_k) w_{ik}^{(2)} h_i(1-h_i) x_j$$

Problem 6



(a) Solve LP feasibility problem for primal SVM

(b)

(P)

$$\min \quad \frac{1}{2}(w_1^2 + w_2^2) + C(\xi_1 + \xi_2 + \xi_3 + \xi_4 + \xi_5 + \xi_6)$$

$$\begin{aligned}
\text{s.t.} \quad & 4w_1 + 4w_2 + b \geq 1 - \xi_1 \\
& 4w_2 + b \geq 1 - \xi_2 \\
& 2w_1 + 2w_2 + b \geq 1 - \xi_3 \\
& b \geq 1 - \xi_4 \\
& 2w_1 + b \leq -1 + \xi_5 \\
& 2w_2 + b \leq -1 + \xi_6 \\
& \xi_1, \xi_2, \dots, \xi_6 \geq 0
\end{aligned}$$

(D)

$$\max \quad \alpha_1+\alpha_2+\alpha_3+\alpha_4+\alpha_5+\alpha_6$$

$$-\frac{1}{2}\Big[16\alpha_1\alpha_2+16\alpha_1\alpha_3+8\alpha_2\alpha_3$$

$$-8\alpha_1\alpha_5-8\alpha_1\alpha_6-8\alpha_2\alpha_5-4\alpha_3\alpha_5-4\alpha_3\alpha_6$$

$$+32\alpha_1^2+16\alpha_2^2+8\alpha_3^2+4\alpha_5^2+4\alpha_6^2\Big]$$

s.t

$$\alpha_1+\alpha_2+\alpha_3+\alpha_4 = \alpha_5+\alpha_6$$

$$0 \le \alpha_1, \ldots, \alpha_6 \le C$$

(C) Only variable that needs non-zero slack is $(0,0)$ and corresponding constraint is

$$-3 \geq 1-\xi_4$$

i.e. $\xi_4 \geq 4$, so. $\xi_4 = 4$ and All other $\xi_i = 0$ and primal value is

$$\frac{1}{2}(1+1)+10\cdot\xi_4 = 41$$

(d) Use SVM solver.

(e) False — see example in Lecture.