Final Exam

Q1: About conditional independent

A: Because a is independent of b, c given d, which means

$$p(a,b,c|d) = p(a|d)p(b,c|d),$$

then

$$\int_c p(a,b,c|d)\mathrm{d}c = \int_c p(a|d)p(b,c|d)\mathrm{d}c,$$

this is $p(a,b|d) = p(a|d)p(b|d)$, which means a is conditional independent of b given d.

Q2: About logistic regression

A: A logistic regression model is defined as

$$p(y|x,w) = \mu(x)^y(1-\mu(x)^{1-y}) \quad \text{with } \mu(x) = sigm(w^Tx+b).$$

Then the logit needed to be calculated is

$$\ln\frac{p(y=1)}{p(y=0)} = \ln\frac{\mu(x)}{1-\mu(x)} = \ln\exp(w^Tx+b) = w^Tx+b,$$

which is a linear function w.r.t. x.

Q3: About quadratic regression

A: Assume the noise distribution is $N(0,\sigma^2)$.

The likelihood is

$$L(w) \propto \exp\left(-\frac{\sum_{i=1}^{n}(y_i - w^2x_i - wx_i)^2}{2\sigma^2}\right).$$

So the log-likelihood is

$$\log L(w) \propto -\frac{1}{2\sigma_2}\sum_{i=1}^{n}(y_i - w^2x_i - wx_i)^2.$$

When the derivative equals to zero, we have

$$\hat{w}^2 + \hat{w} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}.$$

Q4: Model comparison

A: For Model 2, the MLE of w is

$$\hat{w} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}.$$

So they are actually the same model.

Then they have the same fitting accuracy.

Q5: Solution of the Elastic Net Regression.

A: The objective function is $f(w) = RSS(w) + \lambda_1||w||^2 + \lambda_2||w||_1$.

For the RSS term and L_2-norm term, their gradient w.r.t. w_j is $(a_j + \lambda_1)w_j - c_j$, where a_j, c_j are abbreviated constants w.r.t. x, y.

For the L_1-norm term, its sub-gradient is

$$\lambda_1 \cdot \begin{cases} -1 & \text{if } w_j < 0 \\ [-1, 1] & \text{if } w_j = 0 \\ 1 & \text{if } w_j > 0. \end{cases}$$

Then, when the gradient of $f(w) = 0$, it follows that

$$\hat{w}_j = \begin{cases} \dfrac{c_j + \lambda_1}{a_j + \lambda_2} & \text{if } w_j < -\lambda_1 \\ 0 & \text{if } w_j \in [-\lambda_1, \lambda_1] \\ \dfrac{c_j - \lambda_1}{a_j + \lambda_2} & \text{if } w_j > -\lambda_1. \end{cases}$$

Q6: About SVM

A: True. After the training, the founded support vectors wholly determine the decision boundary.

Q7: About GMM

A: The probabilistic membership of these two data are

$$\begin{aligned} p(z_1|x_1) =& \frac{p(x_1|z_1)}{p(x_1|z_1) + p(x_1|z_2)} = \frac{0.4 \times 0.5}{0.2 + 0.12} = \frac{5}{8} \\ p(z_2|x_1) =& 1 - \frac{5}{8} = \frac{4}{8}. \end{aligned}$$

And

$$\begin{aligned} p(z_1|x_2) =& \frac{p(x_2|z_1)}{p(x_2|z_1) + p(x_2|z_2)} = \frac{0.13 \times 0.5}{0.065 + 0.175} = \frac{13}{48} \\ p(z_2|x_2) =& 1 - \frac{13}{48} = \frac{35}{48}. \end{aligned}$$

Q8: About graphical model

A: (1) It is $\{D, I, S, H, L, J\}$, which contains A's and S's parents, co-parents, and children.

(2) (i) True. There is no active path from D to S.

(ii) False. Given $H, D \rightarrow A \rightarrow H$ and $S \rightarrow J \rightarrow H$ consist an active path.

(iii) False. The same reason as the above.

(iv) True. Given A, C is conditional independent with all nodes except D.

Q9: About neural network

A: (1) The square error is

$$l = (y - (c + b\frac{1}{d}\sum_{i=1}^{d} w_i x_i))^2.$$

To minimize the square error by updating w, we calculate the gradient of l w.r.t. w, which is

$$l'(w_i) = 2\left(y - (c + \frac{b}{d}\sum_{i=1}^{d} w_i x_i)\right)\frac{-b}{d}x_i.$$

This is also the update rule to the next w_i.

(2) The according function determined by the neural network is

$$\begin{aligned} y =& c + \frac{b}{d}\sum_{i=1}^{H} w_i \left(\sum_{j=1}^{d} v_{i,j} x_j\right) \\ =& c + \frac{b}{d}\sum_{j=1}^{d}\left(\sum_{i=1}^{H} w_i v_{ij}\right) x_j, \end{aligned}$$

which is a single-layer linear network with weight $\sum w_i v_{ij}$ for input node x_j.