SET A

Q1

Sequence 1
Condition on any one variable and then marginalize on any of the two remaining variables, to get a univariate pdf of the variable that was not conditioned or marginalized over. For Example -
(i) Condition -

$$P_{Y,Z|X}(y,z|X=x) = \frac{P_{X,Y,Z}(x,y,z)}{P_X(x)}$$

(ii) Marginalise -

$$P_{Y|X}(y|X=x) = \int\limits_z P_{Y,Z|X}(y,z|X=x)\,dz$$

Sequence 2
Marginalize on any one of the variable and then condition on any one of the two remaining variables, to get a univariate pdf of the variable that was not conditioned or marginalized over. For Example -
(i) Marginalise -

$$P_{X,Y}(x,y) = \int\limits_z P_{X,Y,Z}(x,y,z)\,dz$$

(ii) Condition -

$$P_{Y|X}(y|X=x) = \frac{P_{X,Y}(x,y)}{P_X(x)}$$

Q2

(i) Less than 1.
Intuition: Data with higher variance is more noisy and unreliable, hence, a better model can be learnt by giving less weightage to the error on this part of the data and a higher weightage to the more reliable data (which has lesser variance).
Mathematical derivation (not required for full credit):
Let there exist n_1 data points with negative x (and variance σ^2) and n_2 data points with non-negative x (and variance $4\sigma^2$).

$$p(y|\boldsymbol{x},\theta,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\frac{(y-f(\boldsymbol{x};\boldsymbol{\theta}))^2}{2\sigma^2}$$

The log-likelihood function is

$$\begin{aligned}
\ell(\boldsymbol{\theta};\sigma) &= \sum_{i=1}^{n_1}\log p\left(y^{(i)}|\boldsymbol{x}^{(i)};\boldsymbol{\theta}\right) + \sum_{j=1}^{n_2}\log p\left(y^{(j)}|\boldsymbol{x}^{(j)};\boldsymbol{\theta}\right) \\
&= \sum_{i=1}^{n_1}\log\left(\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\left(y^{(i)}-f\left(\boldsymbol{x}^{(i)};\boldsymbol{\theta}\right)\right)^2}{2\sigma^2}\right)\right) + \sum_{j=1}^{n_2}\log\left(\frac{1}{\sqrt{2\pi}(2\sigma)}\exp\left(-\frac{\left(y^{(j)}-f\left(\boldsymbol{x}^{(j)};\boldsymbol{\theta}\right)\right)^2}{2\times 4\sigma^2}\right)\right) \\
&= -\frac{1}{2\sigma^2}\sum_{i=1}^{n_1}\left(y^{(i)}-f\left(\boldsymbol{x}^{(i)};\boldsymbol{\theta}\right)\right)^2 - n_1\log\sigma - n_1\log(\sqrt{2\pi}) + \\
&\quad -\frac{1}{2\times 4\sigma^2}\sum_{j=1}^{n_2}\left(y^{(j)}-f\left(\boldsymbol{x}^{(j)};\boldsymbol{\theta}\right)\right)^2 - n_2\log(2\sigma) - n_2\log(\sqrt{2\pi})
\end{aligned}$$

Optimizing θ yields

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n_1} \left(y^{(i)} - f\left(\boldsymbol{x}^{(i)}; \boldsymbol{\theta}\right)\right)^2 + \frac{1}{4} \sum_{j=1}^{n_2} \left(y^{(j)} - f\left(\boldsymbol{x}^{(j)}; \boldsymbol{\theta}\right)\right)^2$$

Hence, $K = \frac{1}{4}$, which is less than 1.

Q3

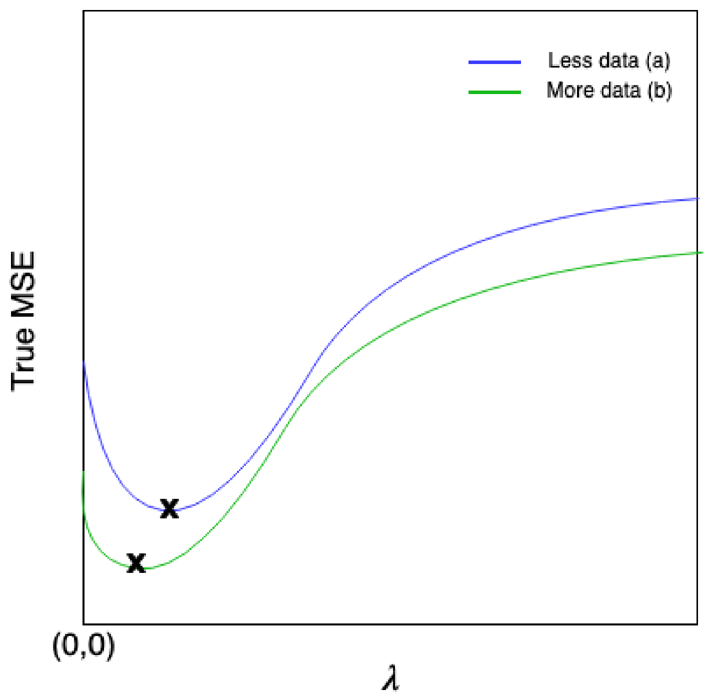

Figure 1: True MSE vs λ

Credits are awarded for shape of the plot and relative positions of the two plots and the two optimal λ.

SET B

Q1

Sequence 1
Condition on any one variable and then marginalize on any of the two remaining variables, to get a univariate pdf of the variable that was not conditioned or marginalized over. For Example -
(i) Condition -

$$P_{Y,Z|X}(y, z|X = x) = \frac{P_{X,Y,Z}(x, y, z)}{P_X(x)}$$

(ii) Marginalise -

$$P_{Y|X}(y|X = x) = \int_z P_{Y,Z|X}(y, z|X = x)\, dz$$

Sequence 2
Marginalize on any one of the variable and then condition on any one of the two remaining variables, to get a univariate pdf of the variable that was not conditioned or marginalized over. For Example -
(i) Marginalise -

$$P_{X,Y}(x, y) = \int_z P_{X,Y,Z}(x, y, z)\, dz$$

(ii) Condition -

$$P_{Y|X}(y|X = x) = \frac{P_{X,Y}(x, y)}{P_X(x)}$$

Q2

(a)

Maximum Likelihood principle is used to determine the values of the parameters of a given parametric probability distribution or probabilistic model that best fits or explains a given dataset.

(b)

NO
MLE is used to estimate the parameters to maximize the likelihood of the available data. It does not guarantee that this model will generalize to unseen data, can lead to over-fitting and there is no mechanism within the maximum likelihood framework to prevent this over-fitting from occurring.

Q3

(a)

The p-value or t-stat is used to check if the corresponding coefficients are sufficiently distinct from 0. The null hypothesis is that a particular coefficient is zero and p-value or t-stat is used to test if the null hypothesis can be rejected. If $t > 2$ (or $p > 0.05$) then the null hypothesis is true and one typically drops the corresponding variable from the model.

(b)

Refer slide 13 and 14 in 2 mlr-short.pdf.
1. Decrease
2. Increase
3. Increase