Problem 1

1. a: formula for the conditional likelihood

NB: $P(Y=1 \mid X_1, \dots X_p) = \dfrac{P(Y=1)\prod_{i=1}^{p} P(X_i \mid Y=1)}{\sum_{j=1}^{2} P(Y=y_j)\prod_{i=1}^{p} P(X_i \mid Y=y_j)}$

LR: $P(Y=1 \mid X_1, \dots X_p) = \dfrac{\exp(w_0 + \sum_{i=1}^{p} w_i X_i)}{1 + \exp(w_0 + \sum_{i=1}^{p} w_i X_i)}$

b. classification rule:

NB: classification rule for $X^{new} = \langle X_1, \dots, X_n \rangle$ is: $Y^{new} \leftarrow \arg\max_{y_k} P(Y=y_k) \prod_i P(X_i^{new} \mid Y=y_k)$

LR: $w_0 + \sum_i w_i X_i \gtrless_{0}^{1} 0$

c. parameters we have to estimate:

For NB: NB for discrete-valued inputs: have to estimate two sets of parameters:

first: $\boxed{\theta_{ijk} \equiv P(X_i = x_{ij} \mid Y=y_k)}$ for each input features X_i, each of its possible values x_{ij}, and each of the possible values y_k of Y.

second: we have to estimate the parameters that define the prior probability over Y:

$\boxed{\pi_k \equiv P(Y=y_k)}$.

NB for continuous Inputs: (for example Gaussian NB) we must estimate the mean and standard deviation of each of these Gaussians

$\boxed{\mu_{ik} \text{ and } \sigma_{ik}^2}$ for each feature X_i and each possible value y_k of Y

For LR: $W \leftarrow \arg\max_W \prod P(Y^l \mid X^l, W)$.

where $\boxed{W = \langle w_0, w_1, \dots w_n \rangle}$ is the vector of parameters to be estimated.