Problem 1

(a) 10%

(b) 10% × 10% = 1%

(C) 10^{-100}

(d) For a fixed p and sample size N, if we want to use training observations that are "near" to a given test sample (near means the training observations are with f range of the test sample in every dimention), then we can only have $N \cdot f^{-p}$ training observations. f^{-p} is very small for large p as $0<f<1$, this means we will have very few training observations 'near' any given test sample

(e) $p=1$: 0.1

$p=2$: $\sqrt{0.1} = 0.316$

$p=100$: $0.1^{\frac{1}{100}} = 0.977$

The answer shows that if we want to use a fixed 'reasonably large' fraction of training observations to make prediction for any give test sample, we will have to use traing observations that are further and further from the test sample in each dimension. These traing observations are no longer "local" to the test sample and the KNN method is expected to perform poorly.

Problem 2

(a) On the training set, QDA will be better. Because it is more flexible. In fact, QDA is fitting the model for Normal distributions that do not necessarily have the same covariance matrices, which contains the situation where all the normals have the same covariance matrices as a subset. Therefore, QDA will fit the training data better.

On the test set, LDA will out perform. Because QDA might over fit the training data, leading to high variance in the $\hat{P}(Y|X)$. However, if n is large, the two model could have similar results.

(b) On the training set, QDA still is better, for same reason as (a).

On the test set, there's not enough information to tell. If the true decision boundary is more close to linear, LDA may out perform, whereas if the true boundary is more close to quadratic, QDA may out perform.

(c) Improve. Because higher n will lead to lower variance for both models. This makes the disadvantage of high variance for QDA to decline gradually. But QDA has lower bias than LDA, therefore, QDA will become better as n increases

(d) True for big training size n. Because if n is big, QDA is less likely to overfit the training data, so it will have an almost linear decision boundary. This makes the test error for LDA and QDA to be similar, then for a certain test set, chances are that QDA will perform better

But for n is small, QDA will overfit the training data and perform worse than LDA

Problem 3.

(a) $\hat{p}(Y=1 \mid X_1=40, X_2=3.5) = \dfrac{1}{1+e^{-(-6+0.05\times 40+1\times 3.5)}} = 0.3775$

(b) $0.5 = \dfrac{1}{1+e^{-(-6+0.05x+3.5)}}$

$e^{2.5-0.05x} = 1.$

$2.5-0.05x = 0$

$x = 50$

$\therefore$ the student needs to study 50 hs

Problem 4

We should use logistic regression. Because the 1-NN method would make no mistake on the training data. Suppose the training and test data both have size N, then the test error rate for 1-NN is: $\dfrac{18\% \times 2N}{N} = 36\%$, which is higher that the test error rate for logistic regression. So we should use logistic regression.