11.1

Parity = True $\rightarrow h(x) = 1$

Parity = False $\rightarrow h(x) = 0$

Since the probability of $P(y=1) = P(y=0) = \frac{1}{2}$

then the distribution is almost 50% 1 and 50% 0.

Hence, Regardless of the parity of the training Set, whether the learning algorithm returns $h(x)=1$ or $h(x)=0$, almost 50% of the time, On the distribution, We will predict the label falsely. $\Rightarrow \boxed{L_D(h) = \frac{1}{2}}$

by parity we mean having even numbers of labels, So, Assume we have a training Set $\underline{\underline{S}}$ which has even numbers of label $\underline{1}$, So it's parity is $\underline{\underline{1}}$.

Apply LOOCV :

Assume the x_i we are leaving out from S has label $\underline{0}$,

$\Rightarrow S - \{x_i\}$ has parity quality $\Rightarrow h(S - \{x_i\}) = 1$

Therefore the error would be $L_v(h(S - \{x_i\})) = 1$

Same is hold for x_i with label $\underline{1}$.

If we omit x_i, $S-\{x_i\}$ would not has the parity quality $\Rightarrow h(S-\{x_i\}) = 0$

So the error would be $L_V(h(S-\{x_i\})) = 1$

because although x_i's label is $\underline{1}$, algorithm will predict $\underline{0}$.

So $|L_D(h) - L_V(h)| = |\frac{1}{2} - 1| = \frac{1}{2}$] ■

11.2

Scenario $\underline{1}$: We have plenty of training examples and $H = \bigcup_{i=1}^{k} H_i$ is not that complex:

$ERM_H(S)$ would be a better choice because it can successfully find the best predictor using the rich training set on the main class H without excessive computations on subclasses.

Scenario 2: We have few training examples and $H = \bigcup_{i=1}^{k} H_i$ is a complex class:

In this case, We can use ensemble models with naive predictors on each class of H_i, then aggregate the results to get an accurate prediction with class H.