SC4000 Machine Learning Tutorial
Ensemble Learning

Question 1: Why is the condition "the base classifiers should do better than a classifier that performs random guessing" necessary for ensemble learning?

Question 2: Suppose we have trained 5 base binary classifiers: f_1, f_2, f_3, f_4 and f_5. Their predictions on a validation dataset are shown in Table 1, where the last column denotes the ground-truth class labels. Which base classifiers would you choose to construct an ensemble learner?

Table 1: Data set for Question 2.

ID	f_1	f_2	f_3	f_4	f_5	Ground Truth
P1	+	+	-	-	+	+
P2	+	+	-	+	-	+
P3	-	-	+	+	-	+
P4	-	-	+	-	+	+
P5	-	-	+	+	-	-
P6	-	-	-	+	+	+
P7	+	+	+	+	-	+
P8	-	+	+	-	+	-
P9	+	+	-	+	+	+
P10	-	-	-	+	-	-