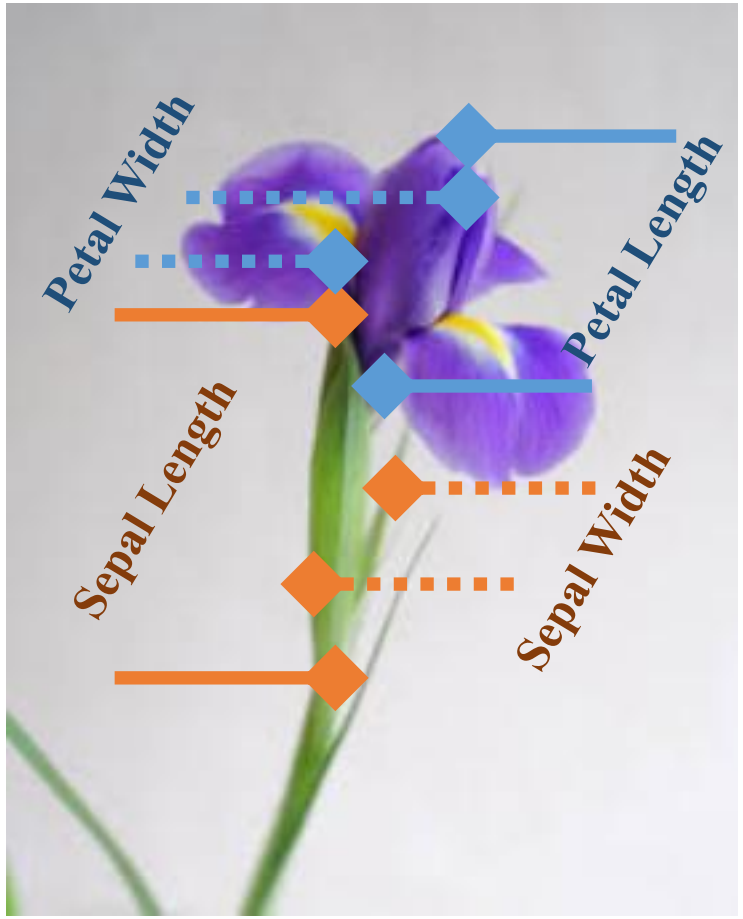
Project 3：鸢尾花数据分类与可视化



特征（Features）：

［Sepal Length，Sepal Width，Petal Length，Petal Width］

x_0 x_1 x_2 x_3

种类（Species）：

［Setosa, , Virginica, Versicolor］

p_0 p_1 p_3

分类器（Classifier）：

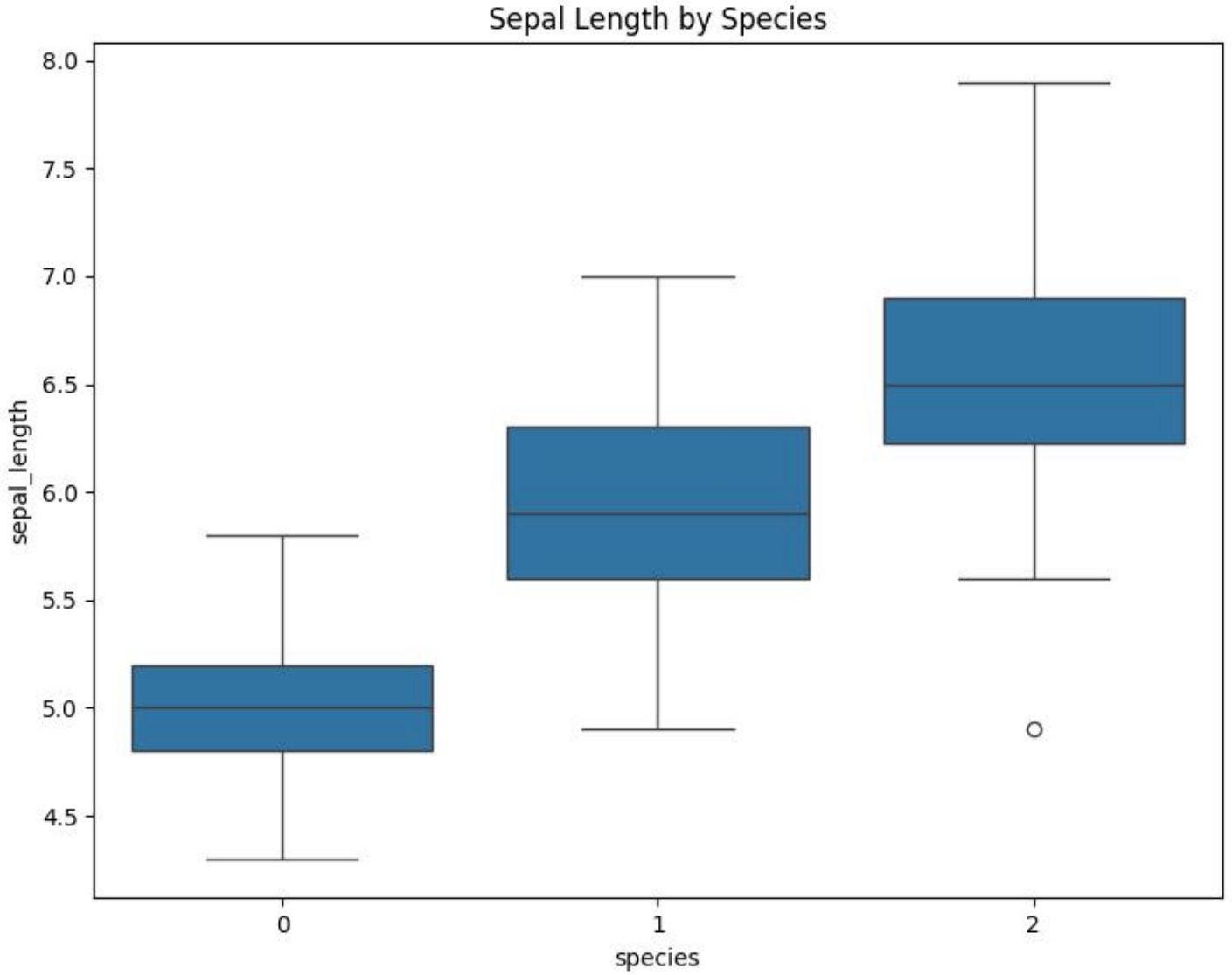
$$p_0, p_1, p_2 = f(x_0, x_1, x_2, x_3)$$

$p_i \geq 0$且$\sum p_i = 1$

Unknow and Solved
by Machine Learning

Project 3：鸢尾花数据分类与可视化

特征（Features）分布可视化



x_0

sns.boxplot:

箱线图，展示数据分布

Project 3：鸢尾花数据分类与可视化

选用合适的分类器（Classifier）

from sklearn.linear_model import LogisticRegression

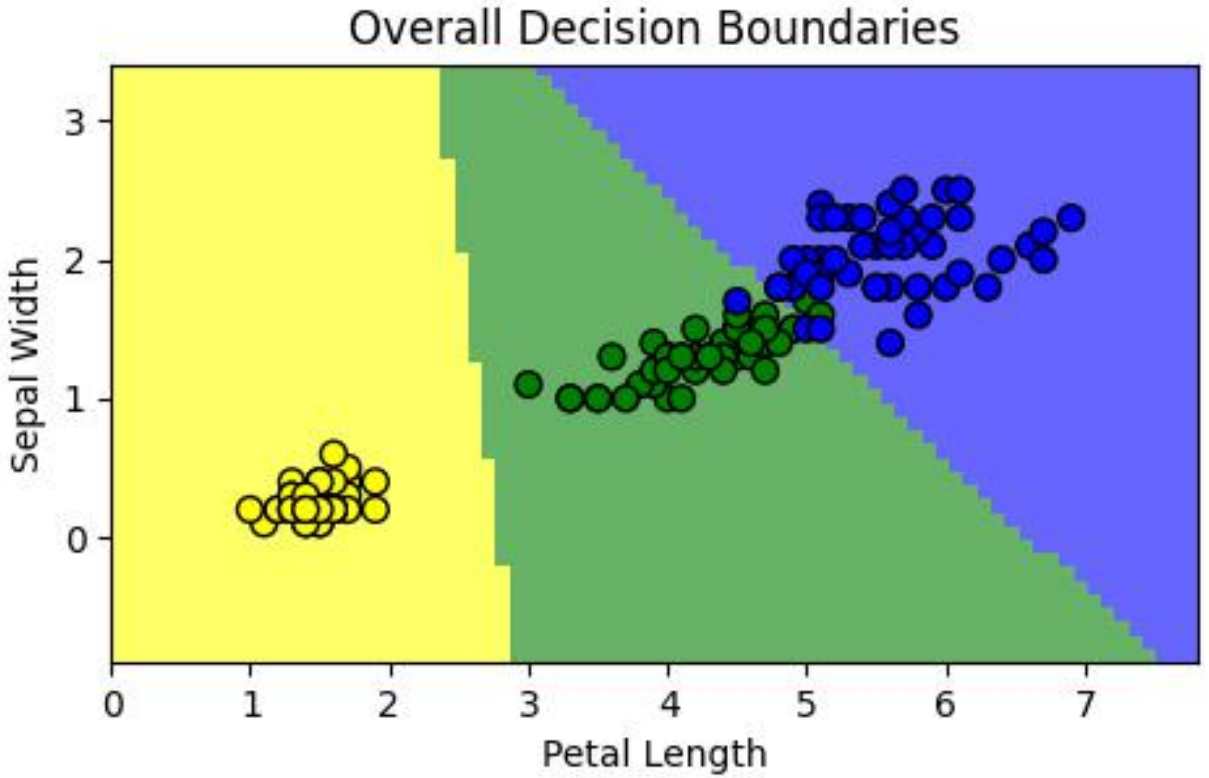
```
classifiers = {
    "Logistic regression\n(C=0.01)": LogisticRegression(C=0.1),
    "Logistic regression\n(C=1)": LogisticRegression(C=100),
    "Gaussian Process": GaussianProcessClassifier(kernel=1.0 * RBF([1.0, 1.0])),
    "Logistic regression\n(RBF features)": make_pipeline(
        Nystroem(kernel="rbf", gamma=5e-1, n_components=50, random_state=1),
        LogisticRegression(C=10),
    ),
    "Gradient Boosting": HistGradientBoostingClassifier(),
    "Logistic regression\n(binned features)": make_pipeline(
        KBinsDiscretizer(n_bins=5, quantile_method="averaged_inverted_cdf"),
        PolynomialFeatures(interaction_only=True),
        LogisticRegression(C=10),
    ),
    "Logistic regression\n(spline features)": make_pipeline(
        SplineTransformer(n_knots=5),
        PolynomialFeatures(interaction_only=True),
        LogisticRegression(C=10),
    ),
}
```

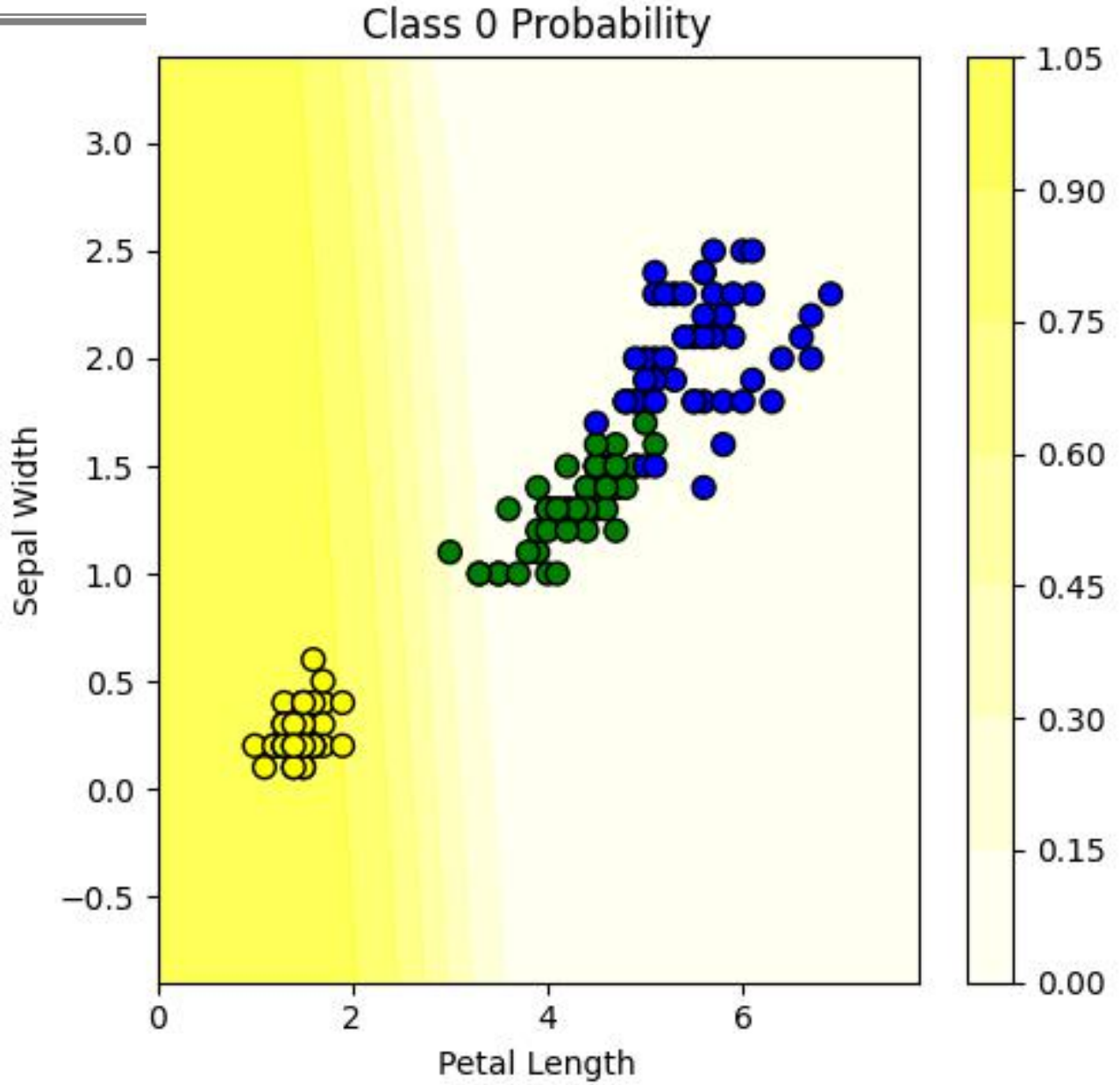
https://scikit-learn.org/stable/auto_examples/classification/plot_classification_probability.html#sphx-

Project 3：鸢尾花数据分类与可视化

$$p_0, p_1, p_2 = f(x_2, x_3)$$

分类边界&概率图可视化

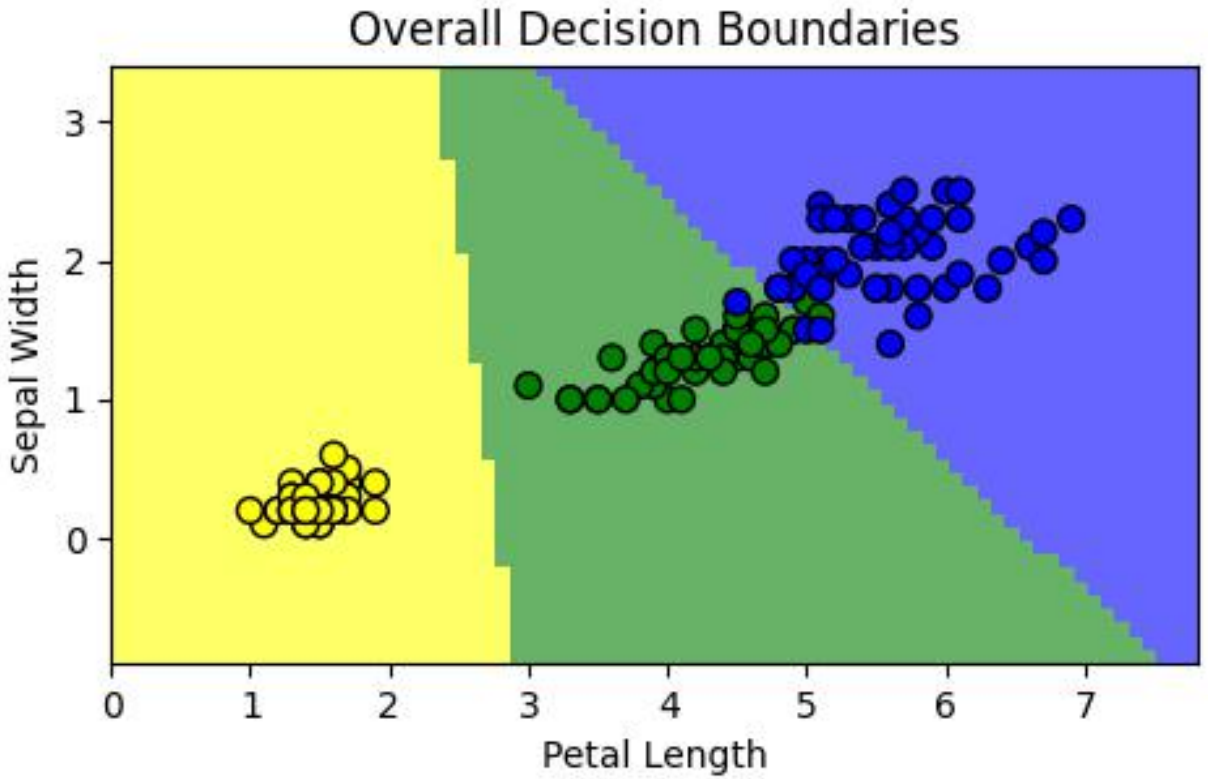


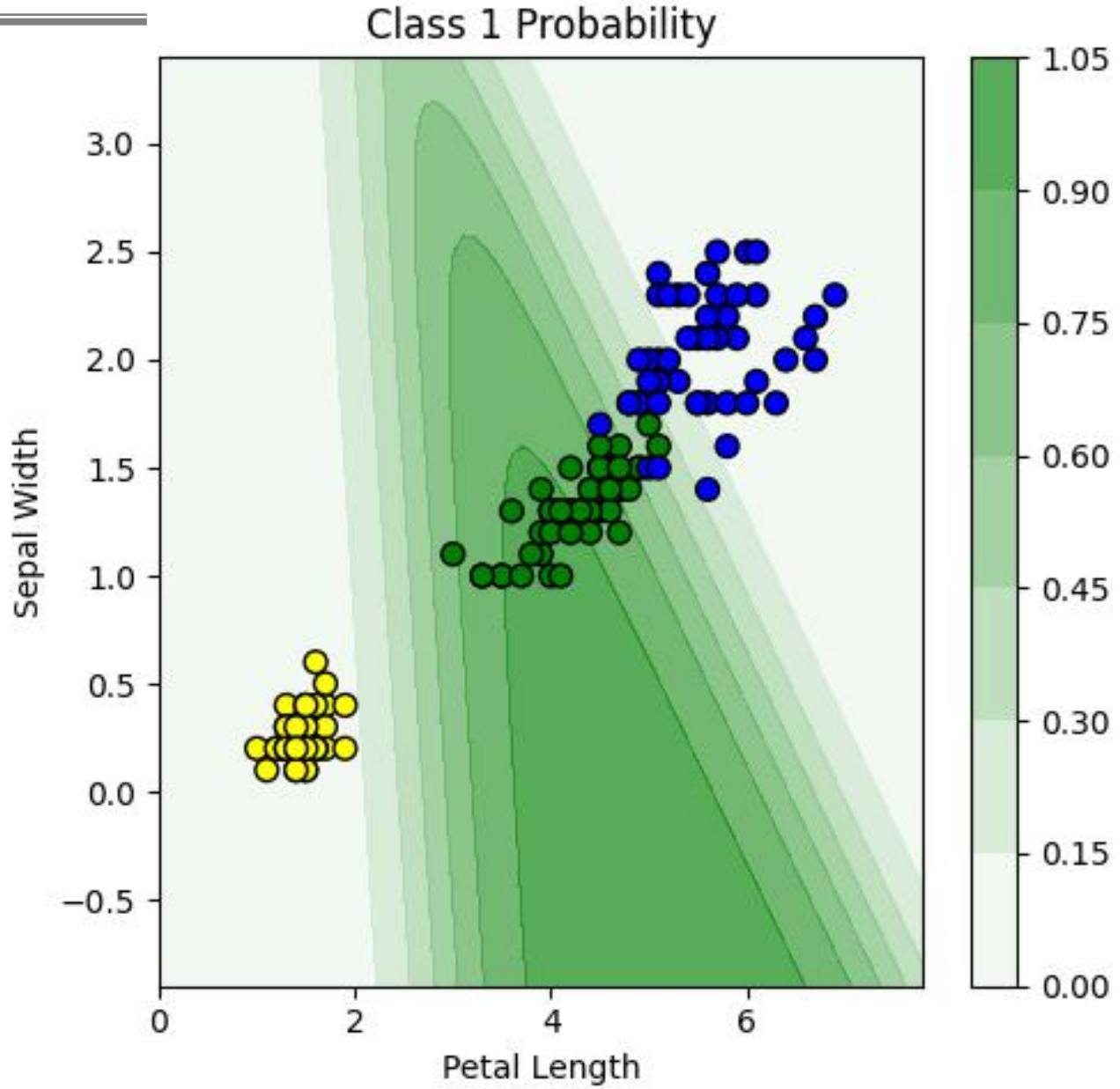


Project 3：鸢尾花数据分类与可视化

$$p_0, p_1, p_2 = f(x_2, x_3)$$

分类边界&概率图可视化

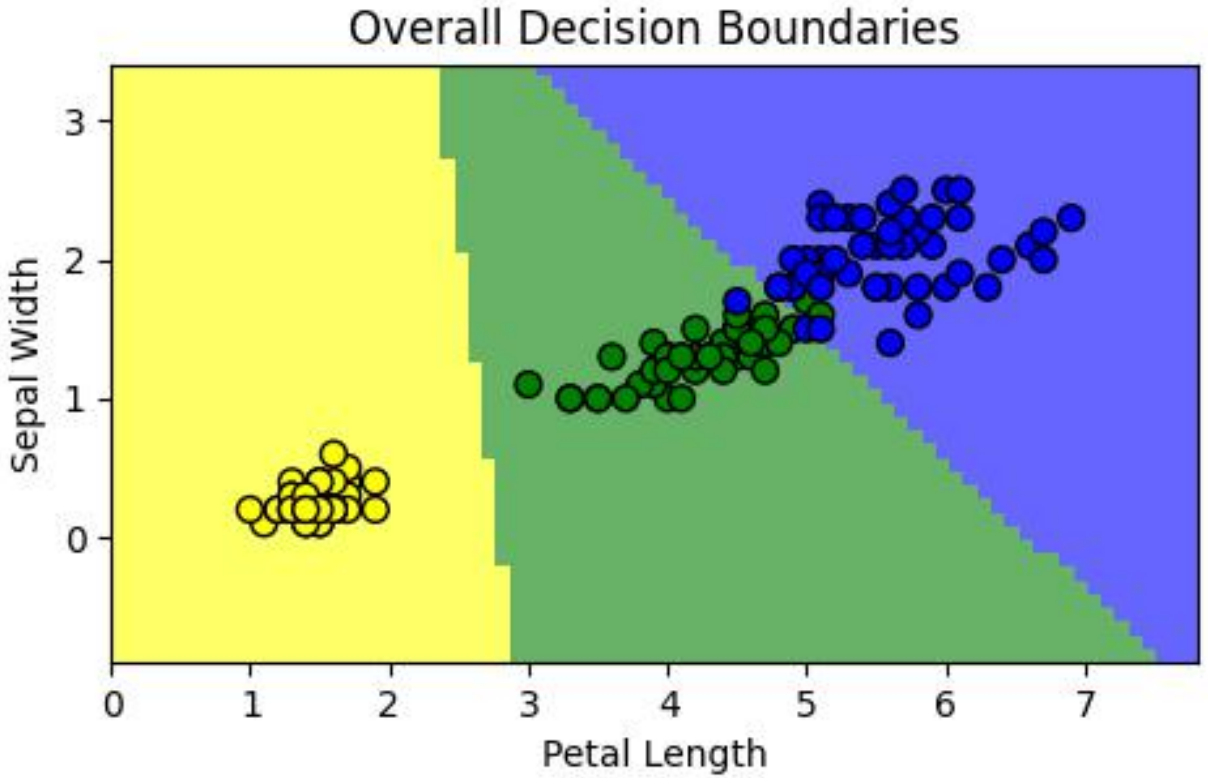


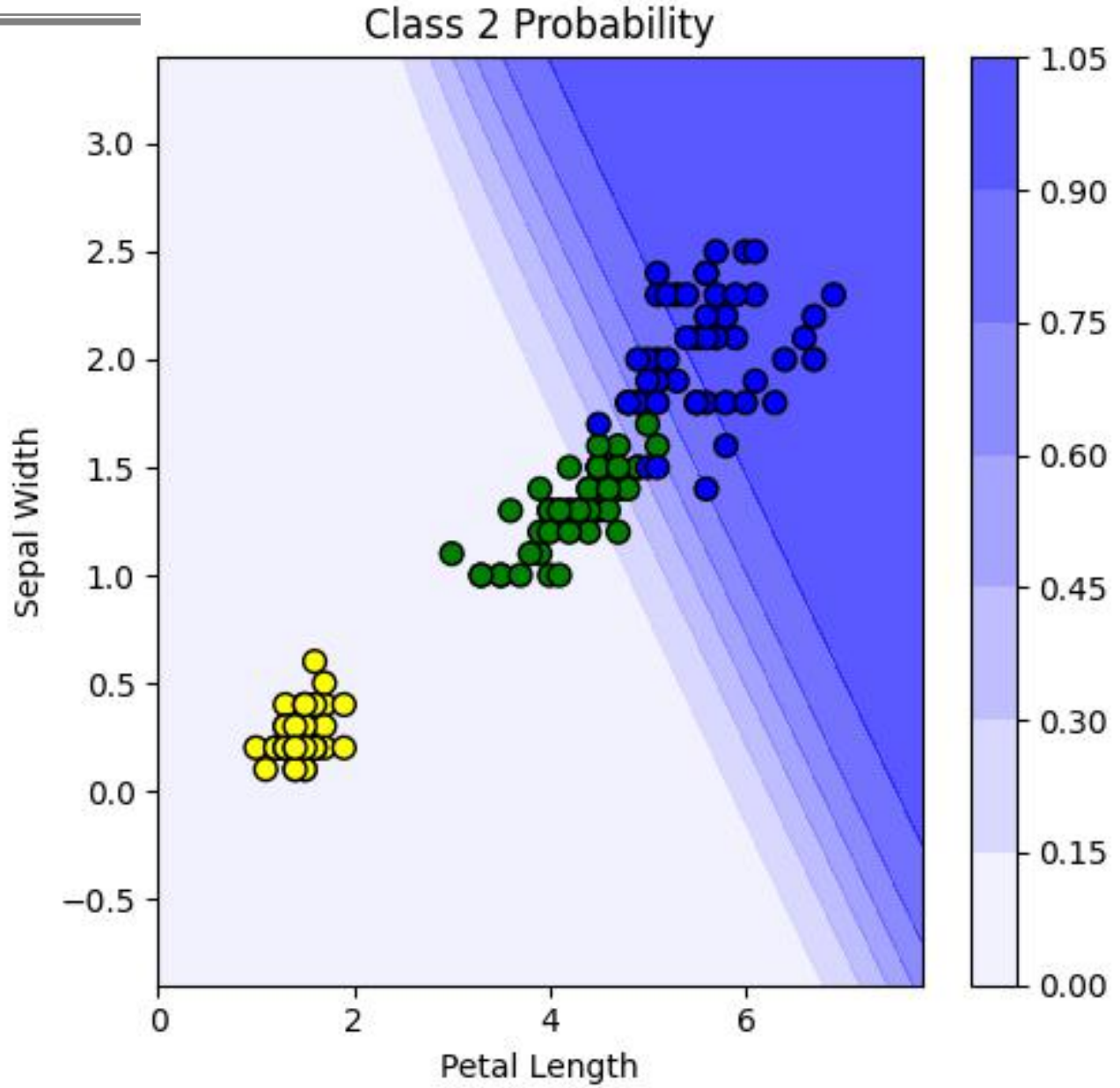


Project 3：鸢尾花数据分类与可视化

$$p_0, p_1, p_2 = f(x_2, x_3)$$

分类边界&概率图可视化



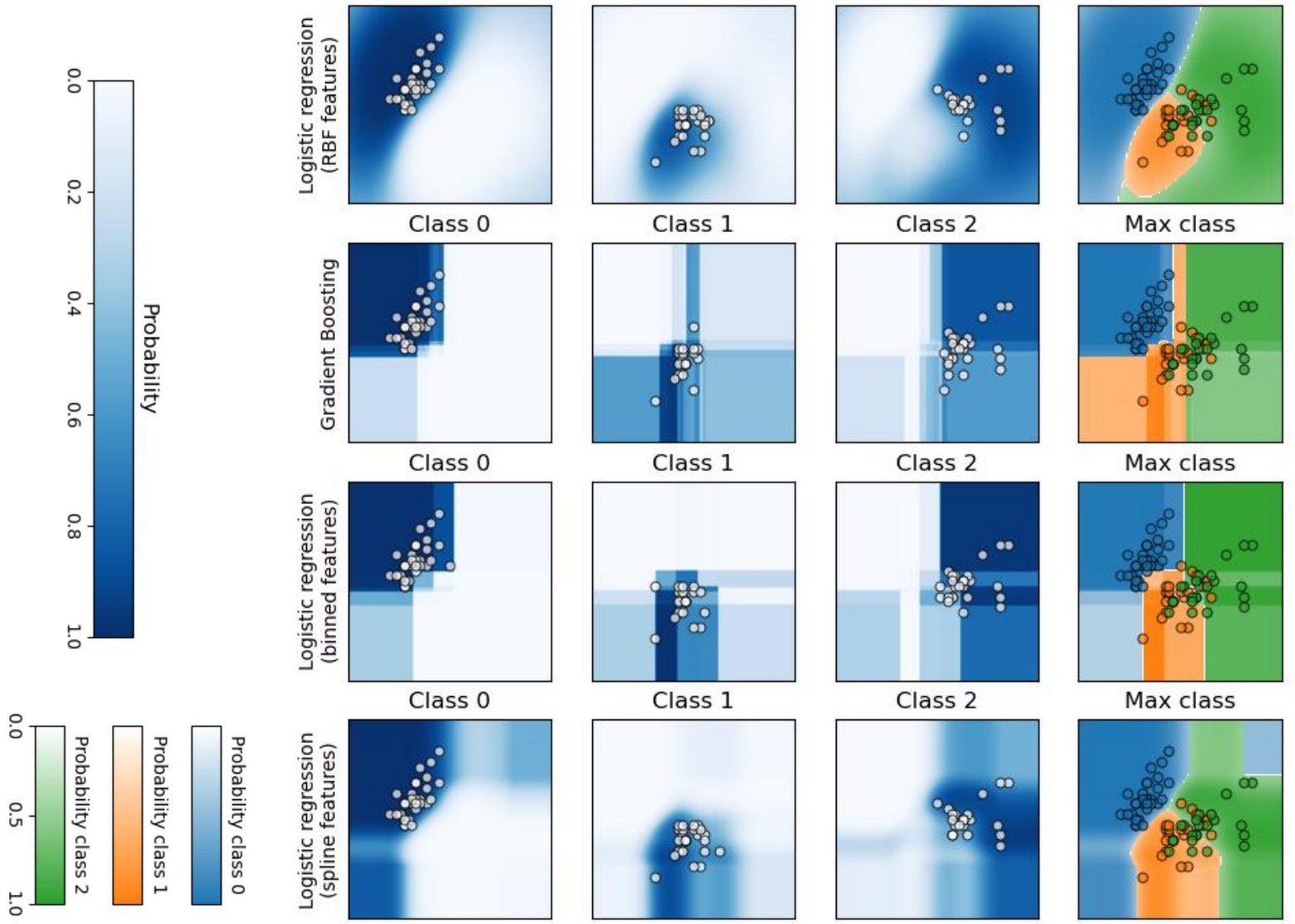


Project 3：鸢尾花数据分类与可视化

任务一：可视化不同分类器的结果

$$p_0, p_1, p_2 = f(x_2, x_3)?$$

三分类/两个特征

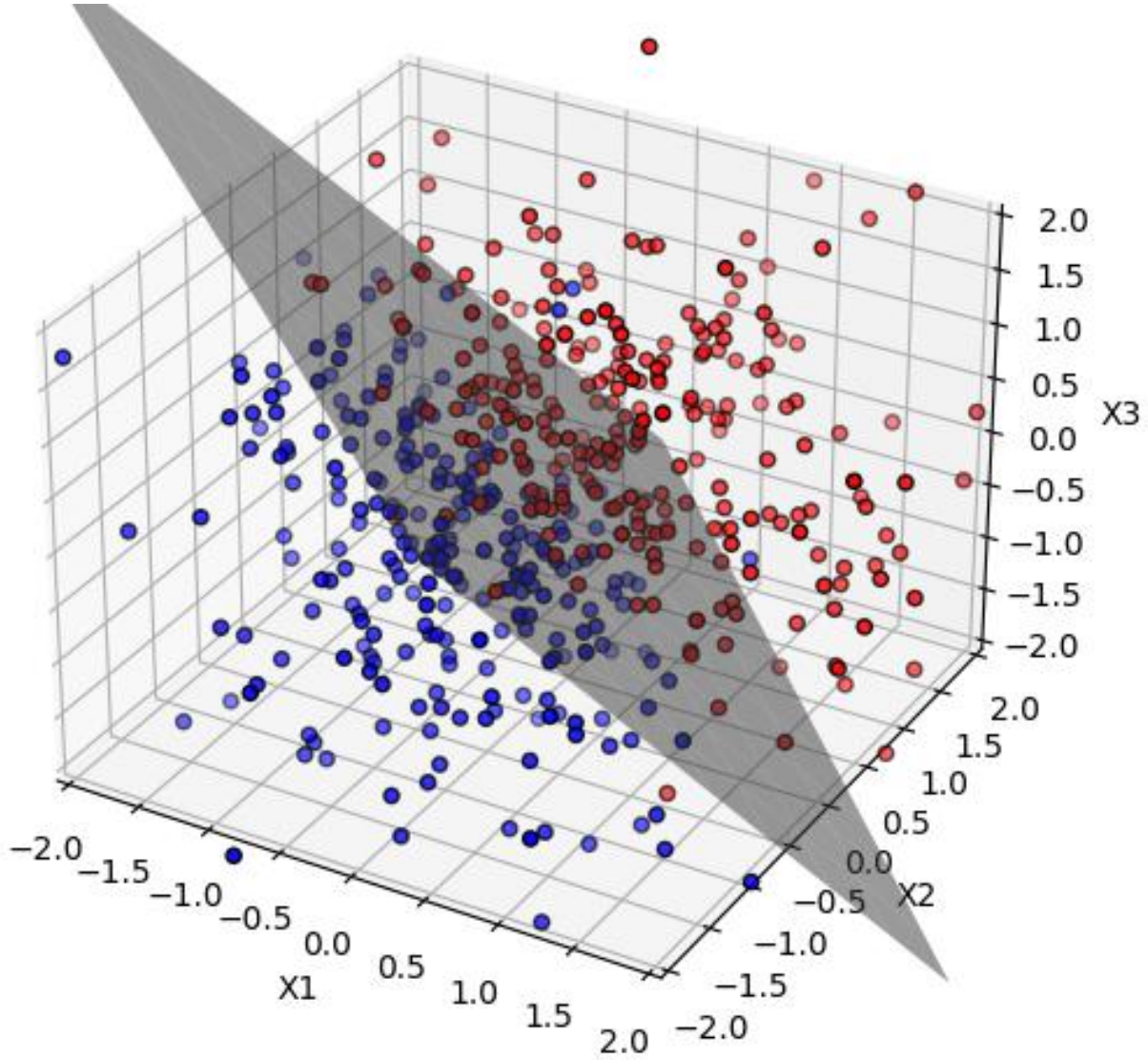


Project 3：鸢尾花数据分类与可视化

任务二：可视化3D Boundary

$$p_0, p_1 = f(x_1, x_2, x_3)?$$

两分类/三个特征

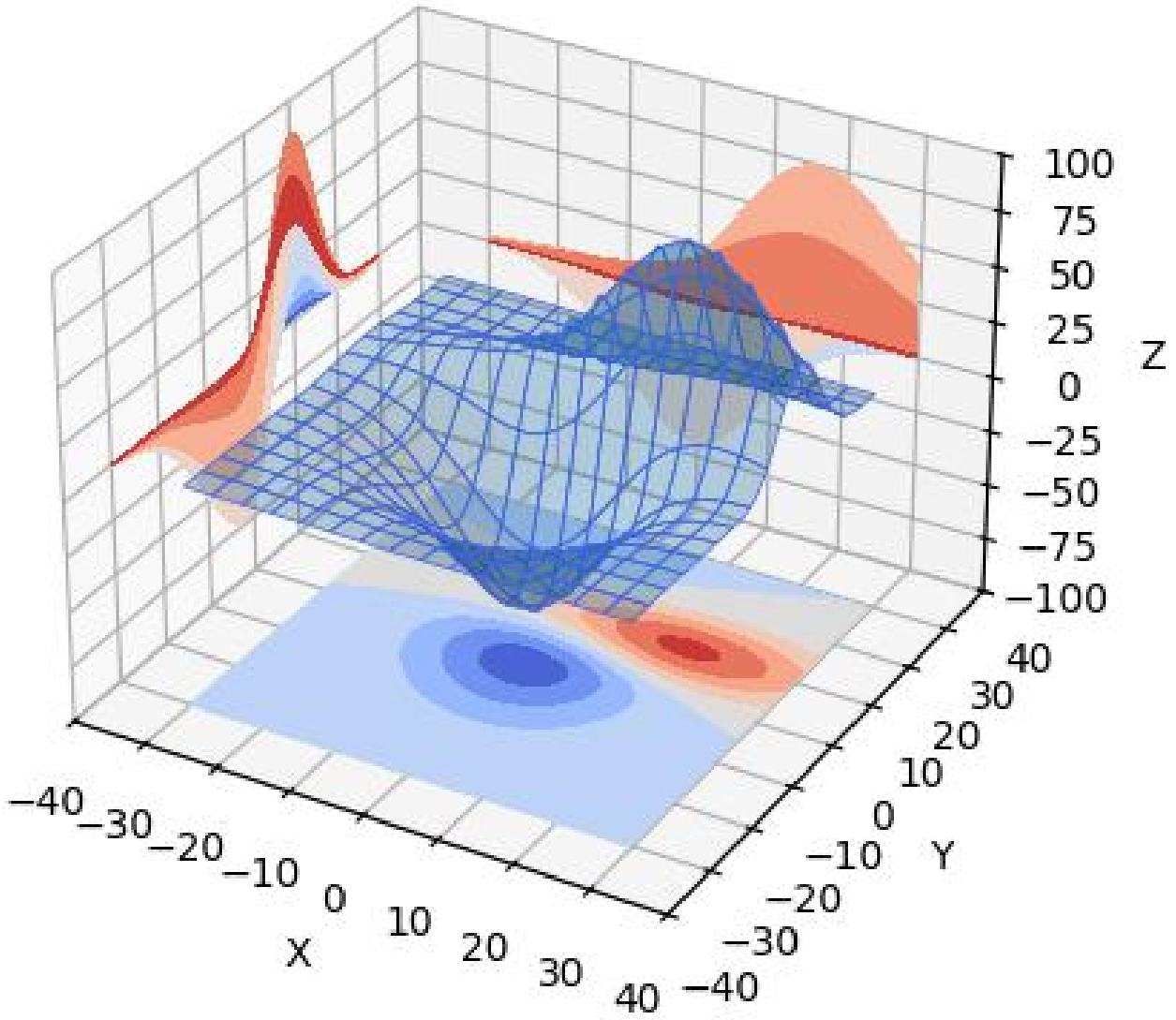


Project 3：鸢尾花数据分类与可视化

任务三：可视化3D Probability Map

$p_0, p_1 = f(x_1, x_2, x_3)$?

两分类/三个特征



Project 3：鸢尾花数据分类与可视化

任务四：能提高分数的任何事情

比如：

$$p_0, p_1, p_2 = f(x_1, x_2, x_3)?$$

3D Boundary + Probability Map