ONLINE MASTERS IN **DATA SCIENCE**

DSC 255 - MACHINE LEARNING FUNDAMENTALS

THE SOFT-MARGIN SUPPORT VECTOR MACHINE

SANJOY DASGUPTA, PROFESSOR

UC San Diego

COMPUTER SCIENCE & ENGINEERING

HALICIOĞLU DATA SCIENCE INSTITUTE

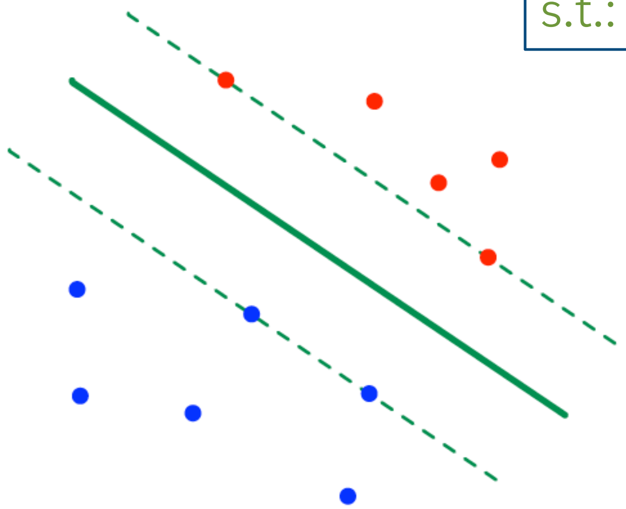
Recall: Maximum-Margin Linear Classifier

Given: training data $\left(x^{(1)}, y^{(1)}\right), \ldots, \left(x^{(n)}, y^{(n)}\right) \in \mathbb{R}^d \times \{-1, +1\}$

Find: the linear separator w that perfectly classifies the data and has maximum margin.

$$\min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \|w\|^2$$

s.t.: $y^{(i)}(w \cdot x^{(i)} + b) \geq 1$ for all $i = 1, 2, \ldots, n$



The solution $w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}$ is a function of just the support vectors.

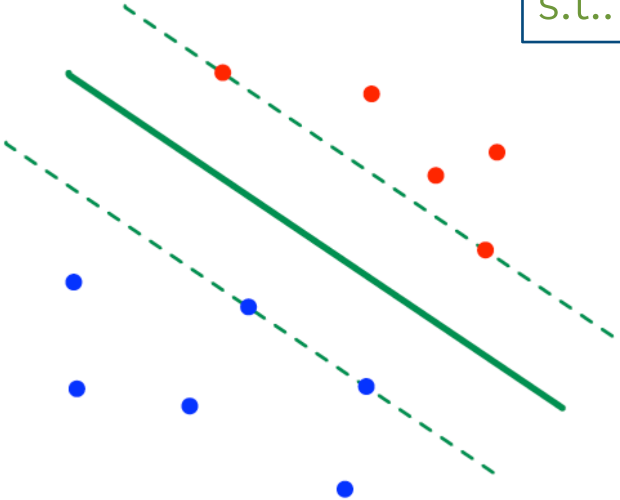
Recall: Maximum-Margin Linear Classifier

Given: training data $\left(x^{(1)}, y^{(1)}\right), \dots, \left(x^{(n)}, y^{(n)}\right) \in \mathbb{R}^d \times \{-1, +1\}$

Find: the linear separator w that perfectly classifies the data and has maximum margin.

$$\min_{w \in \mathbb{R}^d, b \in \mathbb{R}} ||w||^2$$

s.t.: $y^{(i)}(w \cdot x^{(i)} + b) \geq 1$ for all $i = 1, 2, \dots, n$



The solution $w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}$ is a function of just the support vectors.

What if data is not separable?

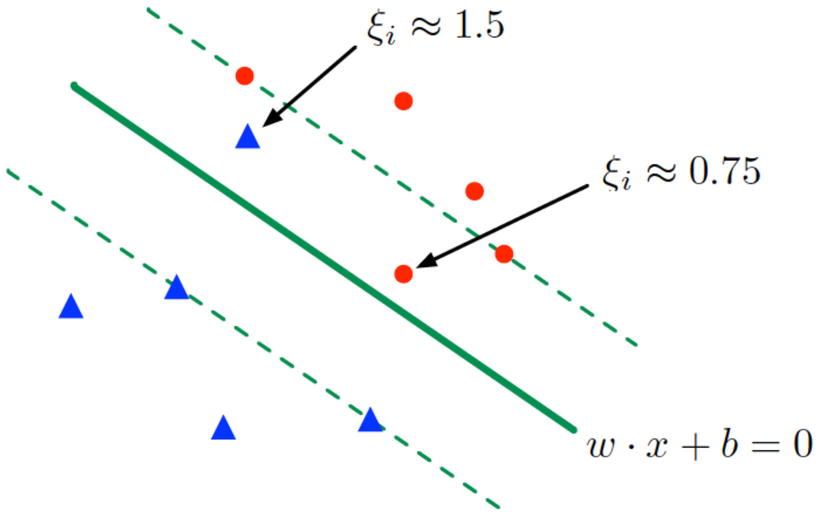
The Non-Separable Case

Allow each data point $x^{(i)}$ some **slack** ξ_i.

$$\min_{w \in \mathbb{R}^d, b \in \mathbb{R}, \xi \in \mathbb{R}^n} \quad \|w\|^2 + C \sum_{i=1}^{n} \xi_i$$

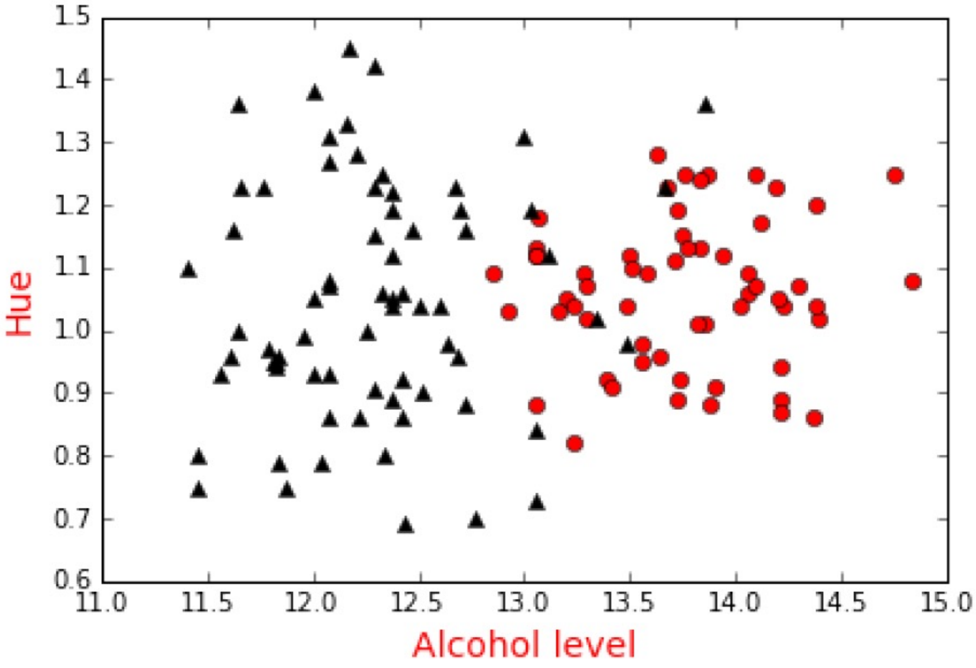
$$\text{s.t.: } y^{(i)}\left(w \cdot x^{(i)} + b\right) \geq 1 - \xi_i \quad \text{for all } i = 1, 2, \dots, n$$

$$\xi \geq 0$$



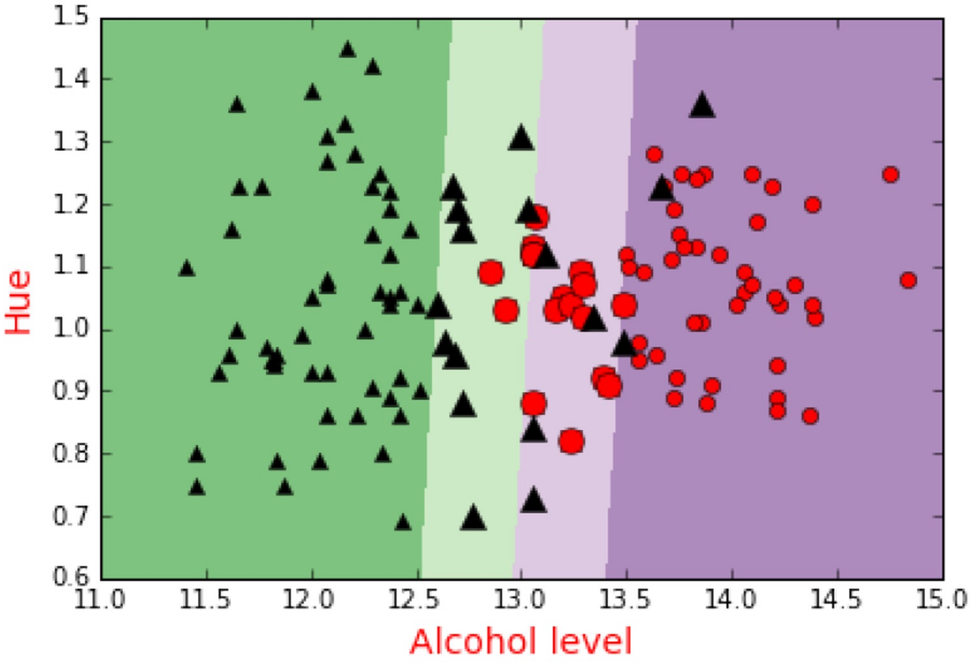
Wine Data Set

Here $C = 1.0$



Wine Data Set

Here $C = 1.0$



The Tradeoff Between Margin & Slack

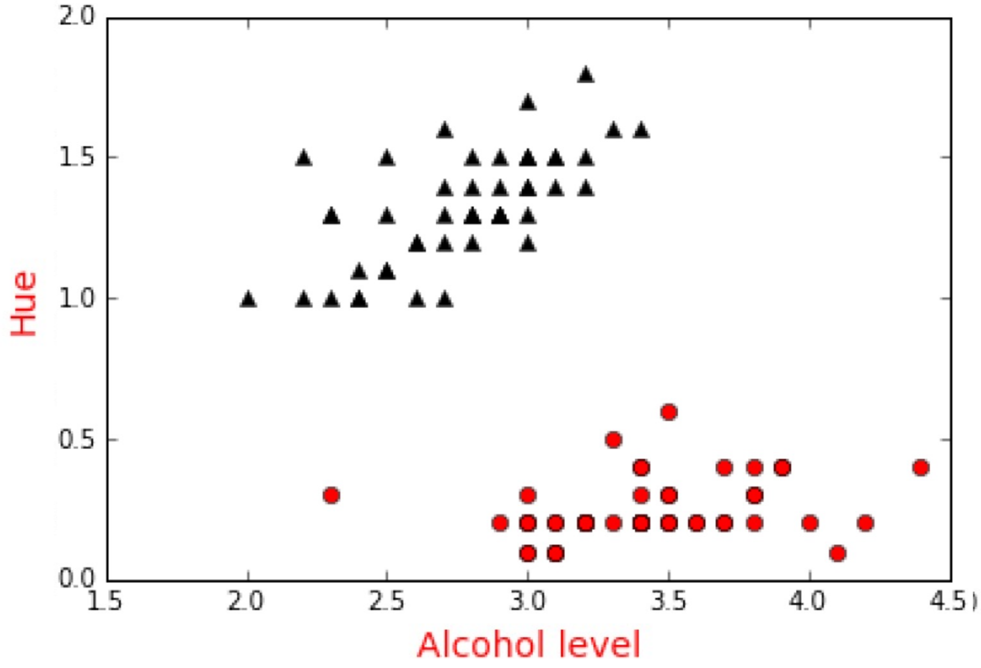
$$\min_{w \in \mathbb{R}^d, b \in \mathbb{R}, \xi \in \mathbb{R}^n} \quad \|w\|^2 + C \sum_{i=1}^{n} \xi_i$$

$$\text{s.t.: } y^{(i)}\left(w \cdot x^{(i)} + b\right) \geq 1 - \xi_i \quad \text{for all } i = 1, 2, \ldots, n$$

$$\xi \geq 0$$

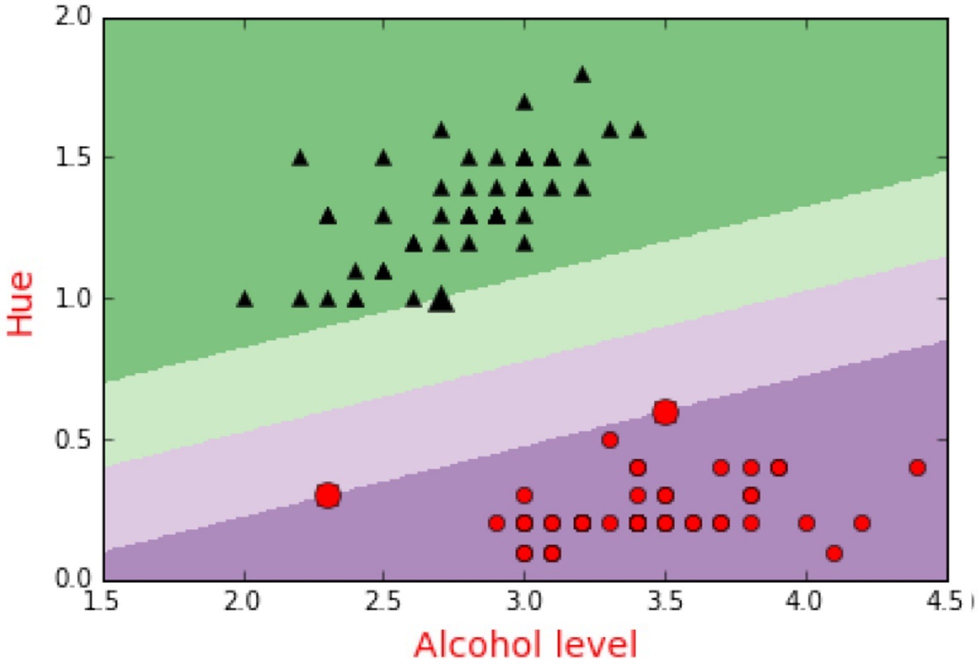
Back to Iris

$C = 10$



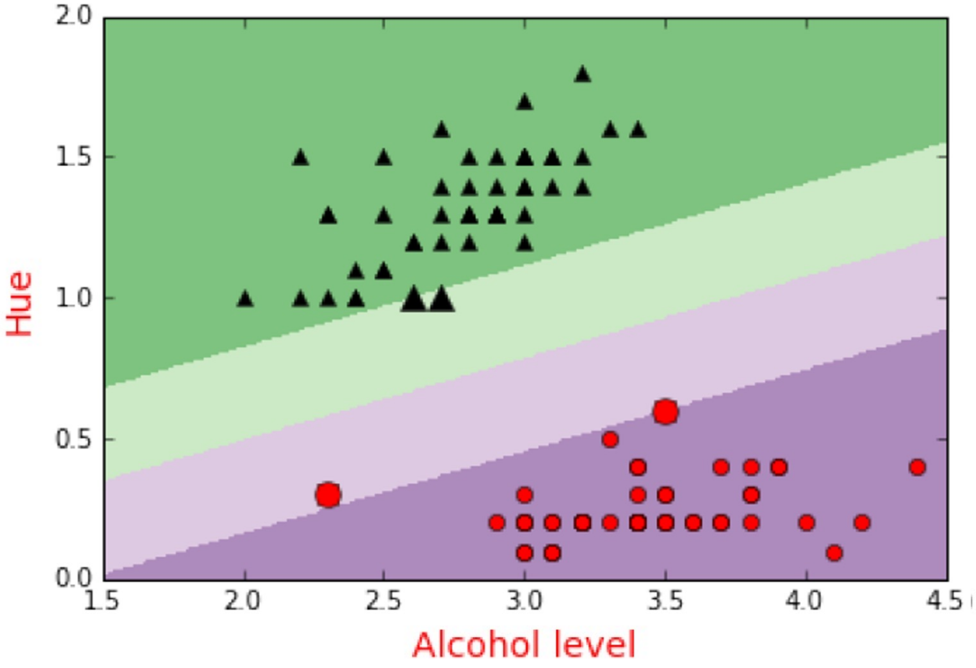
Back to Iris

$C = 10$



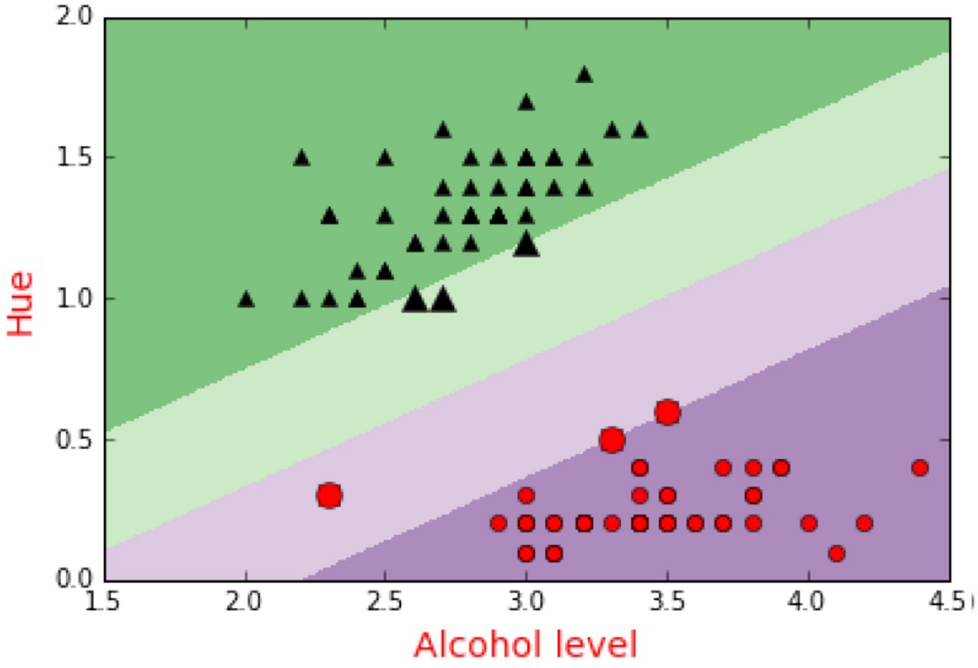
Back to Iris

$C = 3$



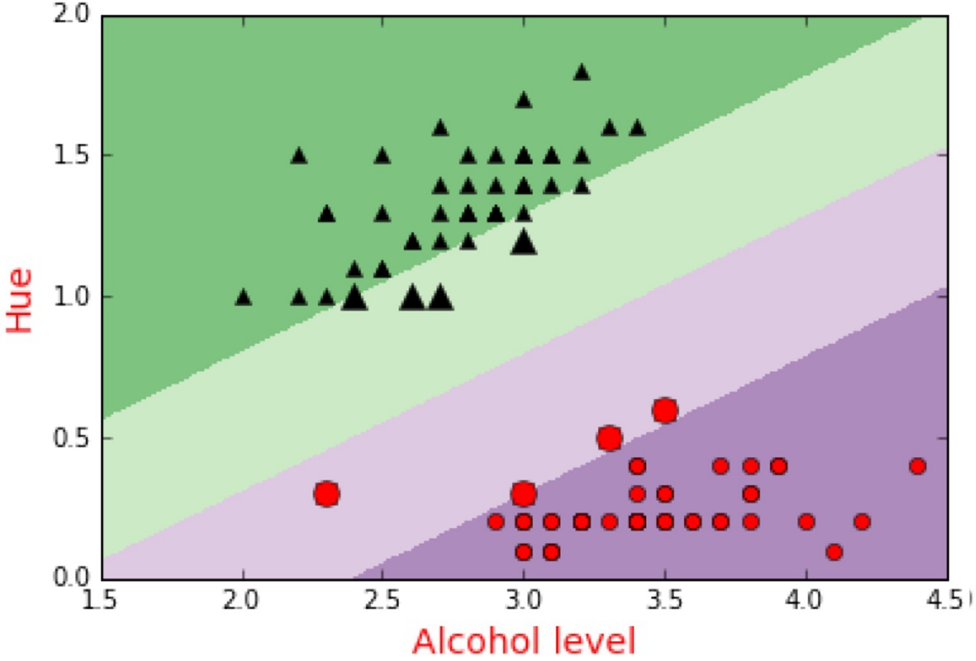
Back to Iris

$C = 2$



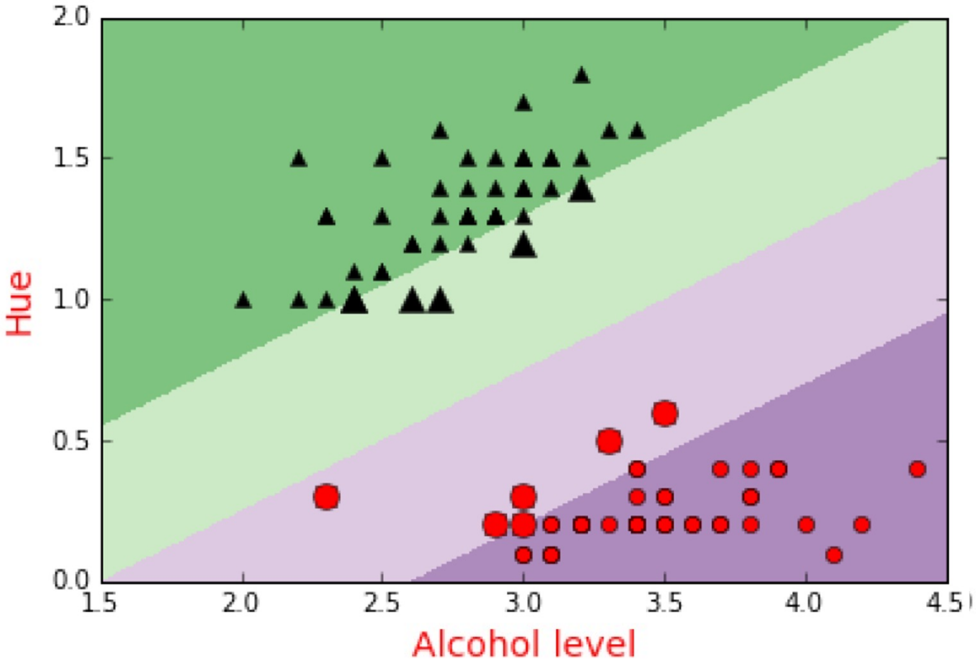
Back to Iris

$C = 1$



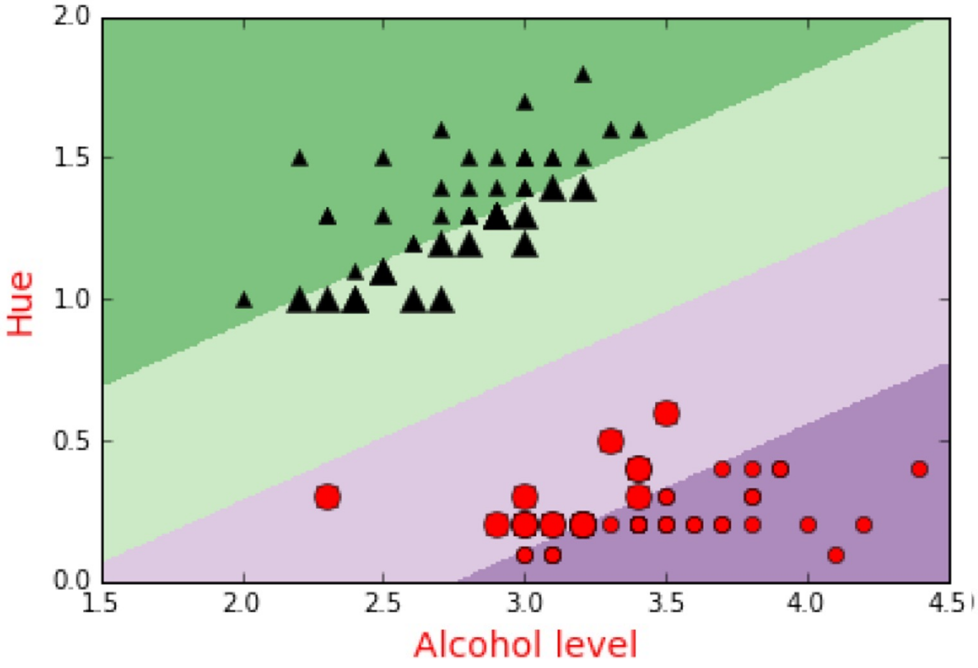
Back to Iris

$C = 0.5$



Back to Iris

$C = 0.1$



Back to Iris

$C = 0.01$