2) SVM with quadratic penalty.

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 + c\sum_i \xi_i^2$$

$$\text{subject to } \quad y_i(w^T x^{(i)} + b) \geq 1 - \xi_i \quad \& \ \forall i \quad \xi_i \geq 0.$$

$$\Rightarrow \ 1 - y_i(w^T x^{(i)} + b) - \xi_i \leq 0 \ , \ -\xi_i \leq 0$$

Here W is a vector $\in \mathbb{R}^{(n)}$ same as $x^{(i)}$

- ξ is a vector $\in \mathbb{R}^{(m)}$ (No. of data samples)

λ_1, λ_2 are vectors corresponding to $1 - y_i(w^T x^{(i)} + b) - \xi_i \leq 0$, $-\xi_i \leq 0$

$$L(W, b, \xi, \lambda_1, \lambda_2) = \frac{1}{2}\|\vec{w}\|^2 + c\sum_{i=1}^{m}\xi_i^2 + \sum_{i=1}^{m}\lambda_1^{(i)}\left(1 - y_i(w^T x^{(i)} + b) - \xi_i\right) - \sum_{i=1}^{m}\lambda_2^{(i)}\xi_i$$

$$\frac{\partial L}{\partial w_1} = w_1 + \sum_{i=1}^{m}\frac{\partial}{\partial w_1}\left(\lambda_1^{(i)}\left(1 - y_i(w^T x^{(i)} + b) - \xi_i\right)\right)$$

$$= w_1 + \sum_{i=1}^{m}\lambda_1^{(i)}\left(-y_i\frac{\partial}{\partial w_1}\left(w_1 x_1^{(i)} + w_2 x_2^{(i)} + \cdots + w_m x_m^{(i)}\right)\right)$$

$$= w_1 + \sum_{i=1}^{m}\lambda_1^{(i)}\left(-y_i x_1^{(i)}\right) = 0 \Rightarrow w_1 = \sum_{i=1}^{m}\lambda_1^{(i)} y_i x_1^{(i)}$$

$$\text{ll}^{\text{ly}} \text{ for other } w_j = \sum_{i=1}^{m}\lambda_1^{(i)} y_i x_j^{(i)} \leftarrow [j - \text{DIMENSION}] \quad \text{— } 1$$

$$\frac{\partial L}{\partial \xi_j} = 2c\,\xi_j + \frac{\partial}{\partial \xi_j}\left(\sum_{i=1}^{m}\lambda_1^{(i)}\left(1 - y_i(w^T x^{(i)} + b) - \xi_i\right)\right) + \frac{\partial}{\partial \xi_j}\left(\sum_{i=1}^{m}\lambda_2^{(i)}\xi_i\right)$$

$$\frac{\partial L}{\partial \xi_j} = 2c\,\xi_j + \frac{\partial}{\partial \xi_j}\left(\sum_{i=1}^{m} \lambda_1^{(i)}\left(1 - y_i\left(w^T x^{(i)} + b\right) - \xi_i\right)\right) - \frac{\partial}{\partial \xi_j}\left(\sum_{i=1}^{m} \lambda_2^{(i)} \xi_i\right)$$

$$\Rightarrow 2c\,\xi_j + \left(-\lambda_1^{(j)}\right) - \lambda_2^{(j)} = 0 \Rightarrow \xi_j = \frac{\lambda_1^{(j)} + \lambda_2^{(j)}}{(2c)} \quad — ②$$

[j — DATA POINTS]

$$\frac{\partial L}{\partial b} = \frac{\partial}{\partial b}\left(\sum_{i=1}^{m} \lambda_1^{(i)}\left(1 - y_i\left(w^T x^{(i)} + b\right) - \xi_i\right)\right) = 0$$

$$= -\sum_{i=1}^{m} \lambda_1^{(i)} y_i = 0 \Rightarrow \lambda_1^{(1)} y_1 + \lambda_1^{(2)} y_2 + \cdots \lambda_1^{(m)} y_m = 0 \quad — ③$$

$$\vec{w} = \sum \lambda_1^{(i)} y_i \vec{x}^{(i)} \quad \text{from } ①$$

$$\Rightarrow \min \frac{1}{2}\|w\|^2 + c\sum_i \xi_i^2 + \sum_{i=1}^{m} \lambda_1^{(i)}\left(1 - y_i\left(w^T x^{(i)} + b\right) - \xi_i\right) - \sum_{i=1}^{m} \lambda_2^{(i)}\left(\xi_i\right)$$

$$\Rightarrow \frac{1}{2}\left(\sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_1^{(i)}\lambda_1^{(j)} y_i y_j \vec{x}^{(i)}\cdot\vec{x}^{(j)}\right) + c\sum_{i=1}^{m}\left(\frac{\lambda_1^{(i)} + \lambda_2^{(i)}}{2c}\right)^2 \quad \left[\lambda_1^{(i)} = -\lambda_2^{(i)} + 2c\,\xi_i\right]$$

$$+ \sum_{i=1}^{m} \lambda_1^{(i)}\left(1 - y_i\left(\sum_{j=1}^{m} \lambda_1^{(j)} y_j \vec{x}^{(j)}\right) x^{(i)} - b\,y_i - \left(\frac{\lambda_1^{(i)} + \lambda_2^{(i)}}{2c}\right)\right)$$

$$- \sum_{i=1}^{m} \lambda_2^{(i)}\left(\frac{\lambda_1^{(i)} + \lambda_2^{(i)}}{2c}\right)$$

$$\Rightarrow \frac{1}{2}\left(\sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_1^{(i)}\lambda_1^{(j)} y_i y_j \vec{x}^{(i)}\cdot\vec{x}^{(j)}\right) + \left(\sum_{i=1}^{m} \lambda_1^{(i)}\right) - b\left(\cancel{\sum \lambda_1^{(i)} y_i}\right)^{0}$$

$$- \sum_{i=1}^{m} \lambda_1^{(i)} y_i \left(\sum_{j=1}^{m} \lambda_1^{(j)} y_j \vec{x}^{(j)}\right) x^{(i)} - \sum_{i=1}^{m}\left(\frac{\lambda_1^{(i)2} + 2\lambda_1^{(i)}\lambda_2^{(i)} + \lambda_2^{(i)2}}{2c}\right)$$

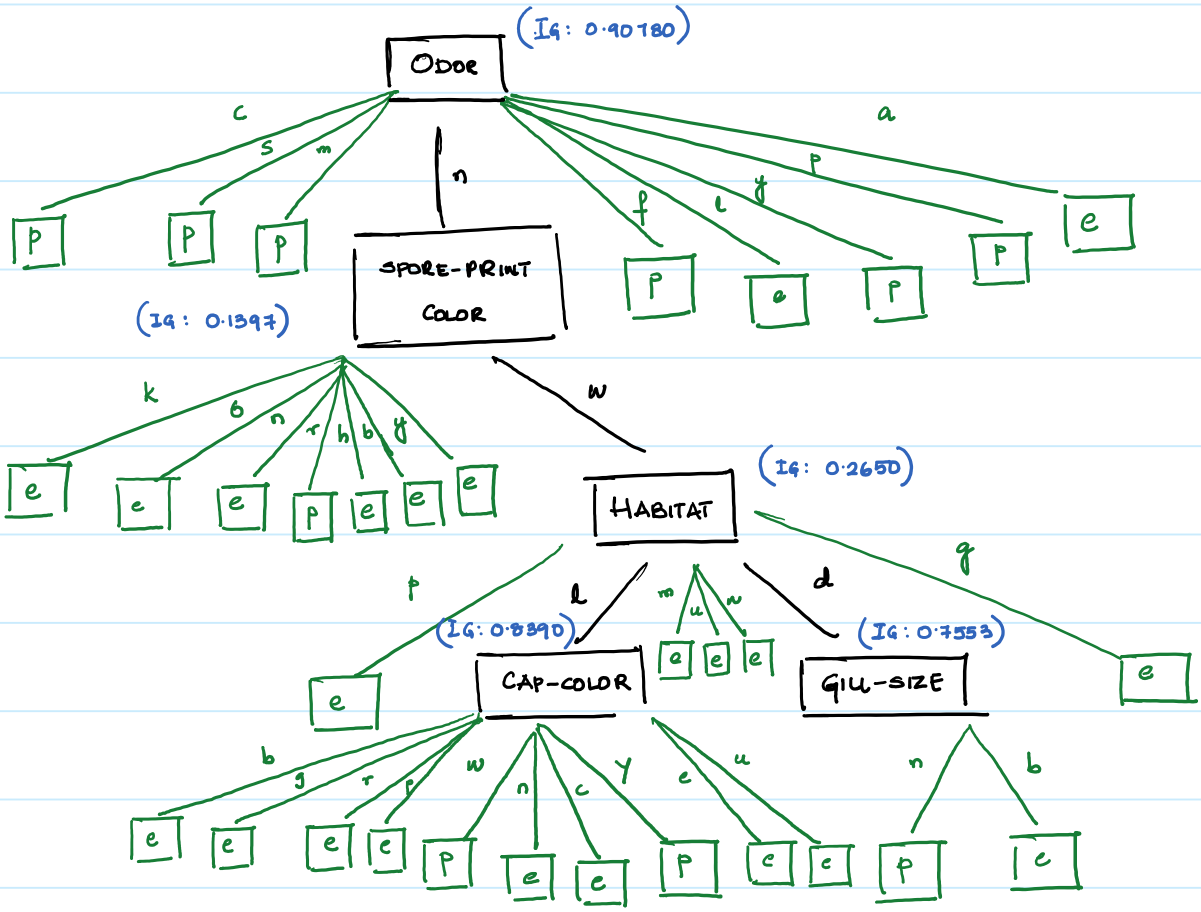
$$= -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_1^{(i)}\lambda_1^{(j)} y_i y_j \underbrace{\vec{x}^{(i)}\vec{x}^{(j)}} + \sum_{i=1}^{m} \lambda_1^{(i)} - \sum\left(\frac{\left(\lambda_1^{(i)} + \lambda_2^{(i)}\right)^2}{4c}\right)$$

Such that $\sum \lambda_1^{(i)} y_i = 0$ & $\lambda_1^{(i)} \geq 0$, $\lambda_2^{(i)} \geq 0$

This does not have an upper bound on $\lambda_1^{(i)}$. KERNEL can be applied.

For this term $\vec{x}^{(i)}\cdot\vec{x}^{(j)}$ can be written as $k\left(\vec{x}^{(i)}, \vec{x}^{(j)}\right)$

3-1) DECISION TREE :



Information gain : ODOR → 0.907

SPORE PRINT COLOR → 0.139

HABITAT → 0.2650

CAP-COLOR → 0.8390

GILL SIZE → 0.7553

3-2) Accuracy of the decision tree on test data is $\boxed{100\%}$

3-3) Considering decision trees of height $\boxed{1}$ i,e one split

$\boxed{\text{YES}}$ it is optimal on the training data. As INFORMATION GAIN is a GREEDY HEURISTIC and for trees of height $\boxed{1}$ its result is optimal.

For multilevel trees, tree produced might not be optimal.