CS461 – RECITATION 11

MACHINE LEARNING PRINCIPLES

Daize Dong

2025-11-24

QUIZ 05

A	B	C	D
–	–	–	–

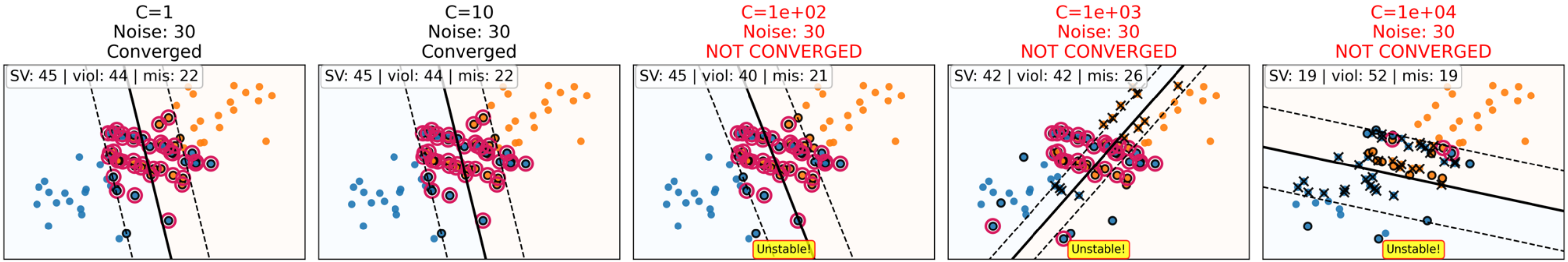
1. What is true about the complexity parameter, C, used with support vector machines?

(a) A very small value of C forces the SVM to use more support vectors.

(b) C is the number of support vectors chosen for each class.

(c) A large value of C makes a simpler model.

(d) All of the above are false.



QUIZ 05

A	B	C	D
–	–	–	–

2. What is true of the kernel trick?

(a) The kernel trick makes large margins in SVMs.

(b) The kernel trick replaces a weight matrix with a function of the training data.

(c) The kernel trick turns perceptrons into SVMs.

(d) None of the above are true.

Dual Objective:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathrm{K}(x_i, x_j)$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \alpha_i y_i = 0 \, , \alpha_i > 0$$

For linear kernel:

$$\mathrm{K}(x, x') = x^T x'$$

$$\Rightarrow \quad \max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

For 2-order polynomial kernel:

$$\mathrm{K}(x, x') = (x^T x' + 1)^2$$

$$\Rightarrow \quad \max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i^T x_j + 1)^2$$

QUIZ 05

A	B	C	D
–	–	–	–

3. When using an SVM, what is true about the weights of the support vectors, α?

(a) If one vector has α = 0.5 and another has α = 0.1, the first is nearer to the decision boundary. α measures how strongly points contribute to the boundary, not *geometric distance.*

(b) The α values must sum to 1.

(c) All α values must be greater than 0.

(d) None of the above are true.

May be misread:
- True if following "support vectors" in title.
- False if α for all points.

- $\alpha_i = 0$: do not affect w, b at all.
- $\alpha_i > 0$: support vectors. These points appear in the model.

  Furthermore, if we take C into account, we have two types of support vectors.

  - $0 < \alpha_i < C$: free support vectors. These points lie exactly on the margin.
  - $\alpha_i = C$: bounded support vectors. These points either lie inside the margin or are misclassified.
  - $\alpha_i > C$: impossible due to the dual constraint.

QUIZ 05

A	B	C	D
–	–	–	–

4. How are SVMs and Neural Networks different?

(a) Only neural networks can work on image data. MNIST

(b) Neural networks are faster to train. SVM: No backpropgation

(c) The kernel trick allows SVMs to learn nonlinear decision boundaries, which NNs cannot. NN learns nonlinear by activation functions.

(d) None of the above are true.

QUIZ 05

A	B	C	D
–	–	–	–

5. When using the kernel trick, how are SVMs and perceptrons different?

(a) SVMs will be slower at inference since they require a complex decision boundary and dense α. Sparse α (those > 0)

(b) The weight matrix of the SVM will not be representable when using a kernel, but it will be when using a perceptron. Obviously representable with linear kernel.

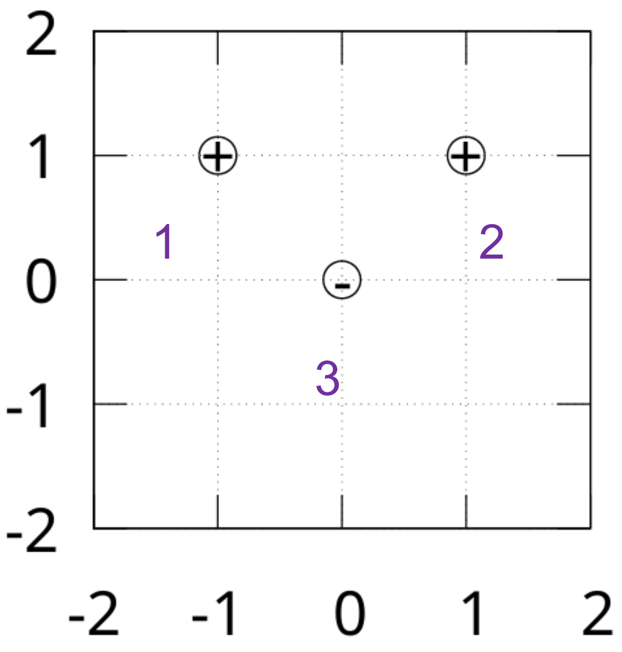
(c) Perceptrons cannot use the kernel trick, but SVMs can. kernel perceptron

(d) SVMs will be faster at inference since they create a decision boundary using a sparse α.

QUIZ 05

Problem 6: Solve for α and w_0 of an SVM.

Part 6.1 (15 points): Given the polynomial kernel $k(x_1, x_2) = (x_1 \cdot x_2)^2$, $X = [(-1,1),(1,1),(0,0)]$ and $Y = (1,1,-1)$, solve for $\alpha_1, \alpha_2, \alpha_3$. Note that the two points from class 1 are symmetrical relative to the third point, so there is a solution where $\alpha_1 = \alpha_2$.



First we get the relationship among all α:

$$\alpha_1 = \alpha_2 \ (1)$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \ \Rightarrow \ \alpha_1 + \alpha_2 - \alpha_3 = 0 \ \Rightarrow \ \alpha_3 = 2\alpha_1 \ (2)$$

Given the polynomial kernel:

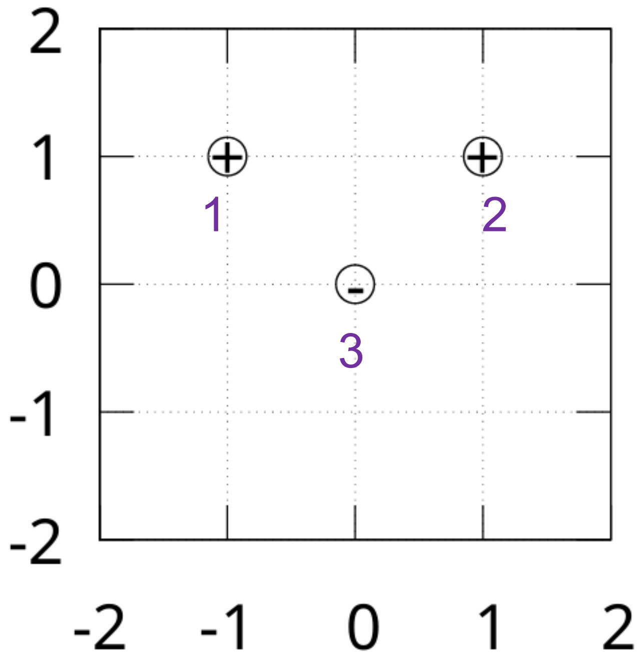
$$\mathrm{K}(x, x') = (x^T x')^2$$

The objective turns to:

$$\sum_{i=1}^{3} \alpha_i - \frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} \alpha_i \alpha_j y_i y_j (x_i^T x_j)^2$$

QUIZ 05

$X = [(-1,1),(1,1),(0,0)]$ and $Y = (1,1,-1)$



$$\sum_{i=1}^{3}\alpha_i - \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\alpha_i\alpha_j y_i y_j (x_i^T x_j)^2$$

$$\Rightarrow (\alpha_1+\alpha_2+\alpha_3) - \frac{1}{2}\begin{pmatrix} 4\alpha_1^2 + 0\alpha_1\alpha_2 + 0\alpha_1\alpha_3 \\ +0\alpha_2\alpha_1 + 4\alpha_2^2 + 0\alpha_2\alpha_3 \\ +0\alpha_3\alpha_1 + 0\alpha_3\alpha_2 + 0\alpha_3^2 \end{pmatrix}$$

$$\Rightarrow \alpha_1+\alpha_2+\alpha_3 - \alpha_1^2 - \alpha_2^2$$

$$\xRightarrow[\alpha_2=\alpha_1 \text{ and } \alpha_3=2\alpha_1]{} -4\alpha_1^2 + 4\alpha_1$$

Take the derivative and it is very easy to deduce:

$$\frac{d}{d\alpha_1}(-4\alpha_1^2 + 4\alpha_1) = -8\alpha_1 + 4 = 0$$

So that:

$$\alpha_1 = \frac{1}{2}$$

$$\alpha_2 = \alpha_1 = \frac{1}{2}$$

$$\alpha_3 = 2\alpha_1 = 1$$

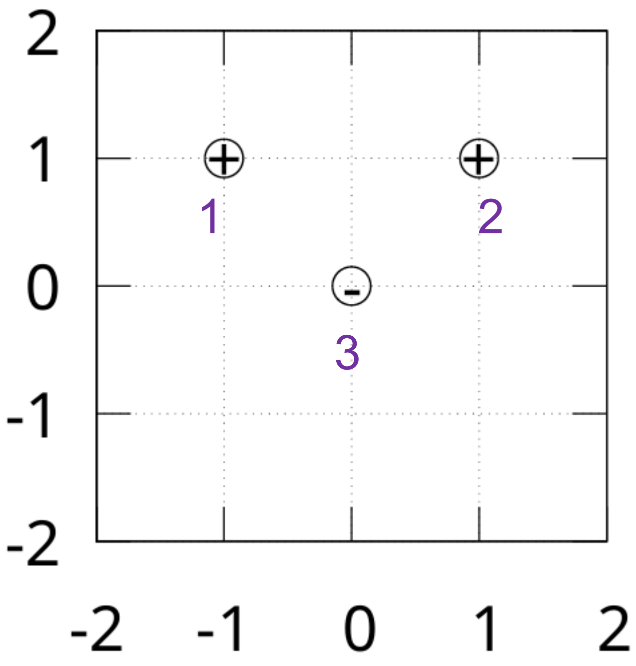
QUIZ 05

$\alpha_1 = \alpha_2 = \frac{1}{2}$

$\alpha_3 = 1$

Part 6.2 (10 points): Using those α values, solve for w_0.

$X = [(-1,1),(1,1),(0,0)]$ and $Y = (1,1,-1)$



$$
\begin{aligned}
w_0 &= y_i - \mathrm{K}(w, x_i) \\
&= y_i - \sum_j^3 \alpha_j y_j \mathrm{K}(x_i, x_j) \\
&= y_1 - \sum_j^3 \alpha_j y_j \mathrm{K}(x_1, x_j) \quad \text{(Take the first point } x_1 \text{ here:)} \\
&= y_1 - \left(\frac{1}{2}({x_1}^T x_1)^2 + \frac{1}{2}({x_1}^T x_2)^2 - 1({x_1}^T x_3)^2\right) \\
&= 1 - \left(\frac{1}{2} * 2^2 + \frac{1}{2} * 0^2 - 1 * 0^2\right) \\
&= 1 - (2 + 0 - 0) \\
&= -1
\end{aligned}
$$