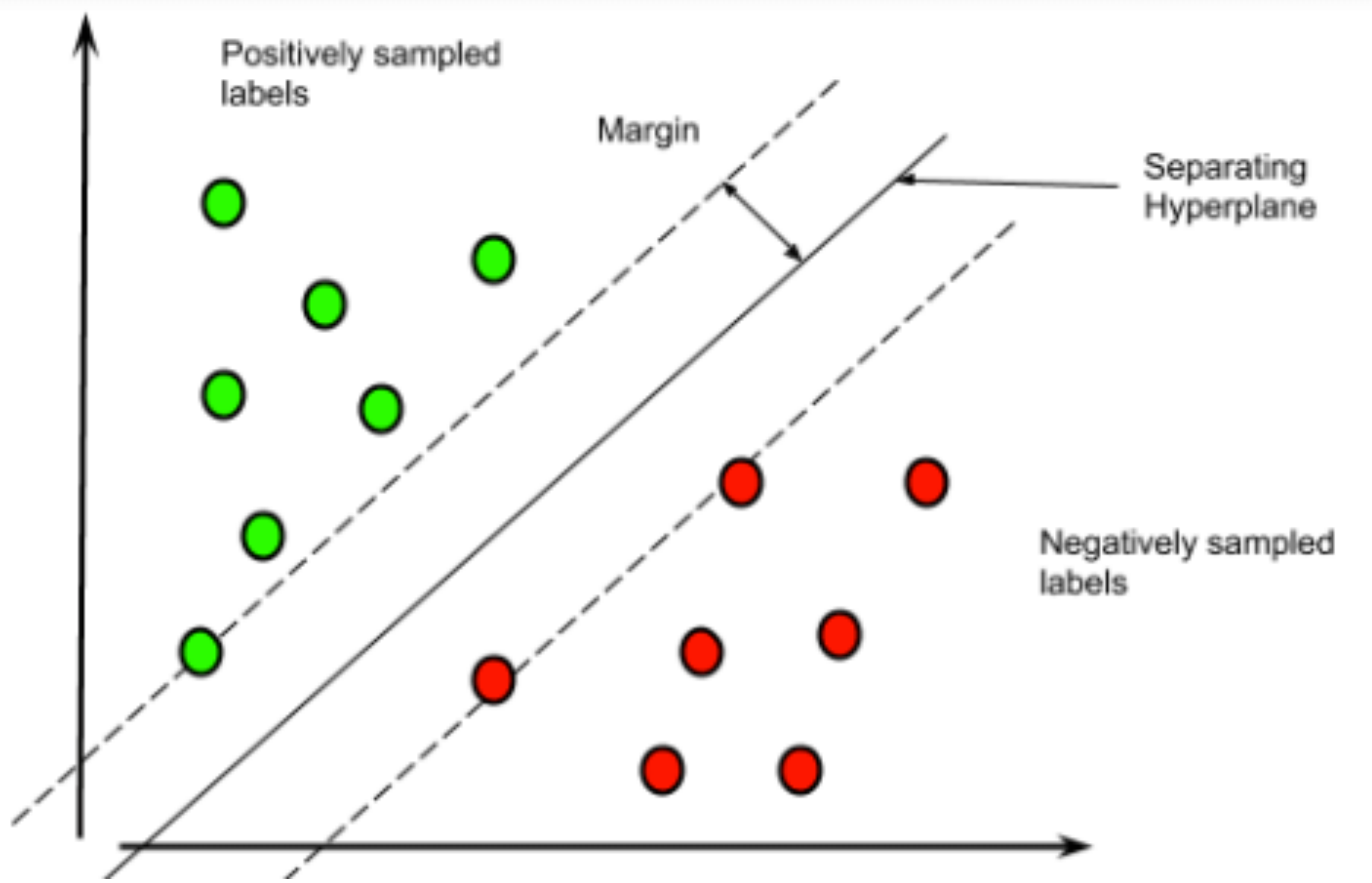


Support Vector Machines: Sequential Minimal Optimisation (SMO)

Leandro L. Minku

Overview

- How to solve the SVM optimisation problem?
  - Breaking the problem into smaller problems.
  - Solving the smaller problems.
  - Heuristics to speed up optimisation.

Dual Representation of Soft Margin Support Vector Machines

$$\text{argmax}_{\mathbf{a}} \tilde{L}(\mathbf{a})$$

$$\tilde{L}(\mathbf{a}) = \sum_{n=1}^{N} a^{(n)} - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} a^{(n)} a^{(m)} y^{(n)} y^{(m)} k(\mathbf{x}^{(n)}, \mathbf{x}^{(m)})$$

Subject to: $0 \leq a^{(n)} \leq C, \forall n \in \{1, \cdots, N\}$

$$\sum_{n=1}^{N} a^{(n)} y^{(n)} = 0$$

This is a constrained optimisation problem.

Solving the Optimisation Problem

- Sequential minimal optimisation is one of the most popular techniques.
    - It breaks the quadratic programming problem into sub-quadratic programming problems that can be solved analytically one at a time.
    - Uses heuristics to decide which of these smaller problems to solve at each step.

Solving the Optimisation Problem

- Sequential minimal optimisation is one of the most popular techniques.
    - It breaks the quadratic programming problem into smaller quadratic programming problems that can be solved analytically one at a time —> a subset of $\mathbf{a}$.
    - Uses heuristics to decide which of these smaller problems to solve at each step —> which subset of $\mathbf{a}$ to solve.

How Many $a^{(n)}$ to Optimise in Each Step?

- Suppose we have $\mathbf{a}$ values for which the constraints are satisfied.
- Suppose we pick $a^{(m)}$ to update, while keeping the others fixed.
- Would we be able to change the value of $a^{(m)}$?

$$\sum_{n=1}^{N} a^{(n)} y^{(n)} = 0 \qquad 0 \le a^{(n)} \le C, \forall n \in \{1, \cdots, N\}$$

- No. Any change in $a^{(m)}$ would result in the summation constraint being violated.

$$a^{(m)} = -y^{(m)} \sum_{n!=m} a^{(n)} y^{(n)}$$

- What is the smallest number of Lagrange multipliers that can be updated in each step such that we can remain satisfying this constraint?

Sequential Minimal Optimisation (SMO)

Initialise $\mathbf{a}$.

Repeat for a maximum number of iterations:

Select a pair of Lagrange multipliers $a^{(i)}$ and $a^{(j)}$ to update next.

Optimise $\tilde{L}(\mathbf{a})$ with respect to $a^{(i)}$ and $a^{(j)}$, while holding all other Lagrange multipliers fixed.

Sequential Minimal Optimisation (SMO)

Initialise $\mathbf{a}$.

Repeat for a maximum number of iterations:

Select a pair of Lagrange multipliers $a^{(i)}$ and $a^{(j)}$ to update next.

Optimise $\tilde{L}(\mathbf{a})$ with respect to $a^{(i)}$ and $a^{(j)}$, while holding all other Lagrange multipliers fixed.

Initialise $\mathbf{a}$

- We can initialise $\mathbf{a}$ to any value that satisfies the constraints.

$$0 \leq a^{(n)} \leq C, \forall n \in \{1, \cdots, N\}$$

$$\sum_{n=1}^{N} a^{(n)} y^{(n)} = 0$$

- The simplest way is to choose $a^{(n)} = 0, \forall n \in \{1, \cdots, N\}$.

Sequential Minimal Optimisation (SMO)

Initialise $\mathbf{a}$.

Repeat for a maximum number of iterations:

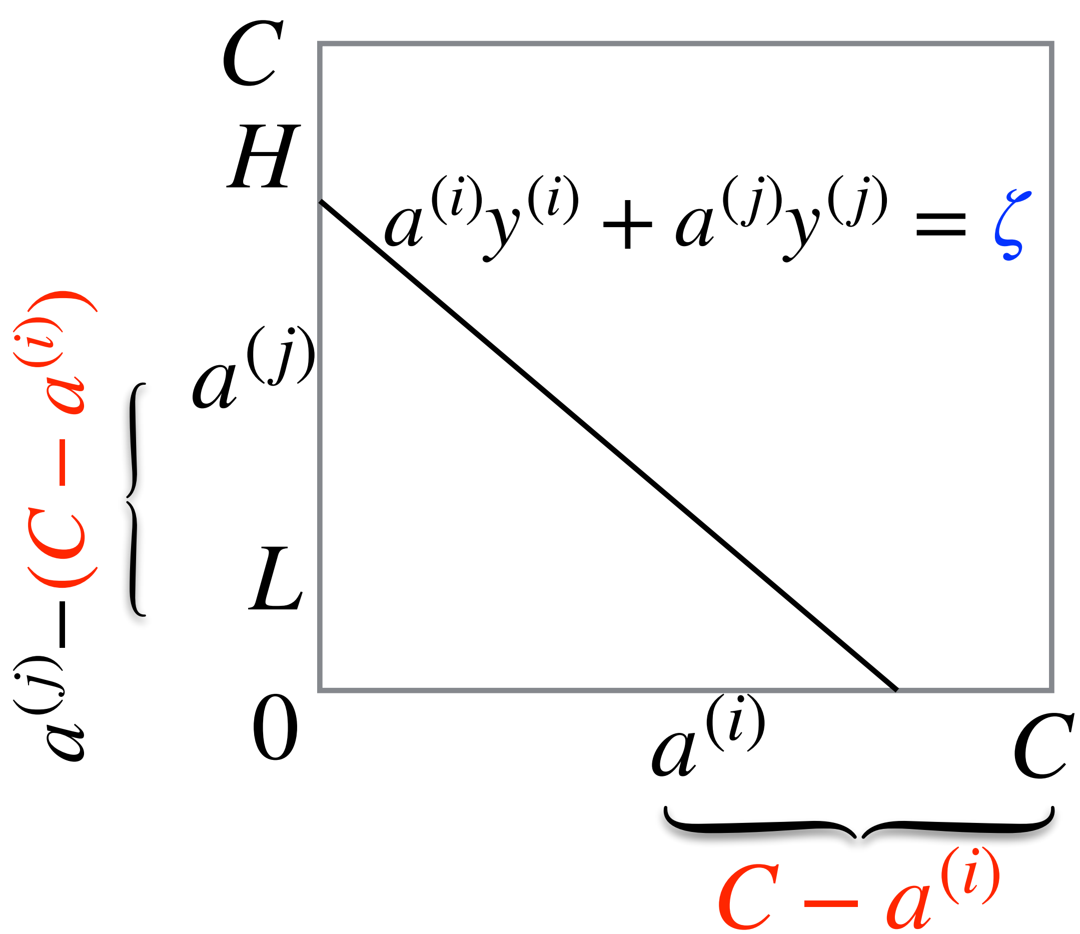
Select a pair of Lagrange multipliers $a^{(i)}$ and $a^{(j)}$ to update next.

Optimise $\tilde{L}(\mathbf{a})$ with respect to $a^{(i)}$ and $a^{(j)}$, while holding all other Lagrange multipliers fixed.

Optimising $L(\tilde{\mathbf{a}})$ With Respect to $a^{(i)}$ and $a^{(j)}$ While Dealing With Constraints

$$0 \leq a^{(n)} \leq C, \ \forall n \in \{1, \cdots, N\}$$

$$\sum_{n=1}^{N} a^{(n)} y^{(n)} = 0 \qquad a^{(i)}y^{(i)} + a^{(j)}y^{(j)} = -\sum_{n \neq i,j} a^{(n)}y^{(n)} \qquad a^{(i)}y^{(i)} + a^{(j)}y^{(j)} = \zeta$$



C

$a^{(i)} + a^{(j)} = \zeta$

0

Assume we are updating $a^{(j)}$ and $y^{(i)} = y^{(j)} = +1$

$$L = \max(0,\ a^{(j)} - (C - a^{(i)}))$$

$$L = \max(0,\ a^{(j)} + a^{(i)} - C)$$

Lower and Higher Possible Values For $a^{(j)}$

- If $y^{(i)} = y^{(j)}$

  $$L = \max(0,\ a^{(j)} + a^{(i)} - C)$$

  $$H = \min(C,\ a^{(j)} + a^{(i)})$$

- If $y^{(i)} \neq y^{(j)}$

  $$L = \max(0,\ a^{(j)} - a^{(i)})$$

  $$H = \min(C,\ C + a^{(j)} - a^{(i)})$$

Writing $a^{(i)}$ as a Function of $a^{(j)}$

$$a^{(i)}y^{(i)} + a^{(j)}y^{(j)} = \zeta$$

$$a^{(i)} = \frac{\zeta - a^{(j)}y^{(j)}}{y^{(i)}}$$

$$\underset{\mathbf{a}}{\operatorname{argmax}}\, \tilde{L}(\mathbf{a}) \longrightarrow \underset{a^{(j)}}{\operatorname{argmax}}\, \tilde{L}(a^{(j)})$$

$$\frac{d\tilde{L}(a^{(j)})}{da^{(j)}} = 0$$

Optimising for $a^{(j)}$

$$a^{(j,new)} = a^{(j)} + \frac{y^{(j)}(E^{(i)} - E^{(j)})}{k(\mathbf{x}^{(i)}, \mathbf{x}^{(i)}) + k(\mathbf{x}^{(j)}, \mathbf{x}^{(j)}) - 2k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})}$$

where the following is the error on training example i:

$$E^{(i)} = h(\mathbf{x}^{(i)}) - y^{(i)}$$

Clipping the Value of $a^{(j)}$

This update rule may lead to violations in the box constraints (not only for $a^{(j)}$ but also for $a^{(i)}$, since $a^{(i)}$ is set based on $a^{(j)}$).

So, the value of $a^{(j)}$ needs to be clipped.

$$a^{(j,new,clipped)} = \begin{cases} H \text{ if } a^{(j,new)} \geq H \\ a^{(j,new)} \text{ if } L < a^{(j,new)} < H \\ L \text{ if } a^{(j,new)} \leq L \end{cases}$$

This will ensure that
the box constraints are satisfied for $a^{(j)}$.

Obtaining $a^{(i)}$

As previously mentioned, $a^{(i)}$ will be set based on $a^{(j)}$:

$$a^{(i,new)}y^{(i)} + a^{(j,new,clipped)}y^{(j)} = \zeta$$

$$a^{(i,new)} = \frac{\zeta - a^{(j,new,clipped)}y^{(j)}}{y^{(i)}}$$

This will ensure that $\sum_{n=1}^{N} a^{(n)}y^{(n)} = 0$ is satisfied
(and the box constraint for $a^{(i)}$ too).

Sequential Minimal Optimisation (SMO)

Initialise $\mathbf{a}$.

Repeat for a maximum number of iterations:

Select a pair of Lagrange multipliers $a^{(i)}$ and $a^{(j)}$ to update next.

Optimise $\tilde{L}(\mathbf{a})$ with respect to $a^{(i)}$ and $a^{(j)}$, while holding all other Lagrange multipliers fixed.

Karush-Kuhn-Tucker (KKT) Conditions

- For soft margin SVM, they lead to the following, $\forall n \in \{1,\ldots,N\}$:
  - $0 < a^{(n)} < C$ iif $y^{(n)}h(\mathbf{x}^{(n)}) = 1$ Support vectors on the margin
  - $a^{(n)} = C$ iff $y^{(n)}h(\mathbf{x}^{(n)}) \leq 1$ Support vectors on or violating the margin
  - $a^{(n)} = 0$ iff $y^{(n)}h(\mathbf{x}^{(n)}) \geq 1$ Not support vectors

Selecting Pair $a^{(i)}$ and $a^{(j)}$ To Optimise Next

- Selection is done heuristically.
- We first select an $a^{(i)}$ and then an $a^{(j)}$.
- The Lagrange multiplier $a^{(i)}$ is selected by alternating between the following two strategies:
  - Selected randomly among those corresponding to training examples that violate the KKT conditions with a certain margin of error e around the value of $y^{(n)}h(\mathbf{x}^{(n)})$.
  - Selected randomly among those corresponding to training examples that violate the KKT conditions given the margin e and $0 < a^{(i)} < C$ (support vectors on the margin).

Selecting Pair $a^{(i)}$ and $a^{(j)}$ To Optimise Next

- The Lagrange multiplier $a^{(j)}$ is selected by trying the following strategies in the following order, until a positive improvement in the objective is observed:
  - Pick the $a^{(j)}$ associated to the example that will obtain the largest change in the $a^{(j)}$ value (which would hopefully result in a large increase in the objective). Largest step size approximated based on $|E^{(i)} - E^{(j)}|$.

$$a^{(j,new)} = a^{(j)} + \frac{y^{(j)}(E^{(i)} - E^{(j)})}{k(\mathbf{x}^{(i)}, \mathbf{x}^{(i)}) + k(\mathbf{x}^{(j)}, \mathbf{x}^{(j)}) - 2k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})}$$

  - Pick each $0 < a^{(j)} < C$ in turn.
  - Look through the entire training set.
  - Replace $a^{(i)}$ and try again.

Summary: Sequential Minimal Optimisation (SMO)

Initialise $\mathbf{a}$.

Repeat for a maximum number of iterations:

Select a pair of Lagrange multipliers $a^{(i)}$ and $a^{(j)}$ to update next.

Optimise $\tilde{L}(\mathbf{a})$ with respect to $a^{(i)}$ and $a^{(j)}$, while holding all other Lagrange multipliers fixed.

Next Week

- Catch up week.
- No new content.
- Schedule:
  - Monday: revision lecture.
  - Thursday: usual tutorial.
  - Friday: tutorial with TAs.

Assignment

- Types of questions:
  - Multiple choice
  - Multiple answers
  - Enter number

- You may need to demonstrate a detailed understanding of how approaches work

- You may need to do calculations

Deadline

The deadline for this assignment is a strict deadline. Late submissions will get mark of zero. You have only one attempt at the assignment, i.e., once you submit it, you will not be able to edit it again.

If your ability to work on the assignment is affected by illness or other external factors then you should contact the welfare team by emailing cswelfare@contacts.bham.ac.uk.

You can start the assignment at any time from the "available from" date below until the "due date". However, once you start, you will have a **maximum of 1 hour** to submit your assignment, except for students with Reasonable Adjustment Plans (RAP), where the time limit will be defined by the RAP. Your assignment will be automatically submitted at the due time. For example, if you start your assignment less than 1 hour before the due time, you will have less than 1 hour to complete your assignment.

Calculations and Numbers with Decimal Cases

Feel free to make your computations using as many decimal cases as you wish but no less than 4 decimal cases. If necessary, feel free to round your computations to four decimal cases. For example, a number of 15.123456 could be rounded as 15.1235.

Multiple Answer Questions (Tickbox Questions)

Beware that to calculate scores for this kind of questions, Canvas divides the total points possible by the amount of correct answers for that question. This amount is **awarded** for every correct answer selected and **deducted** for every incorrect answer selected. Please bear this in mind when selecting answers. If you would like to read a more detailed scoring overview with some examples, this can be found here under "select grading option".