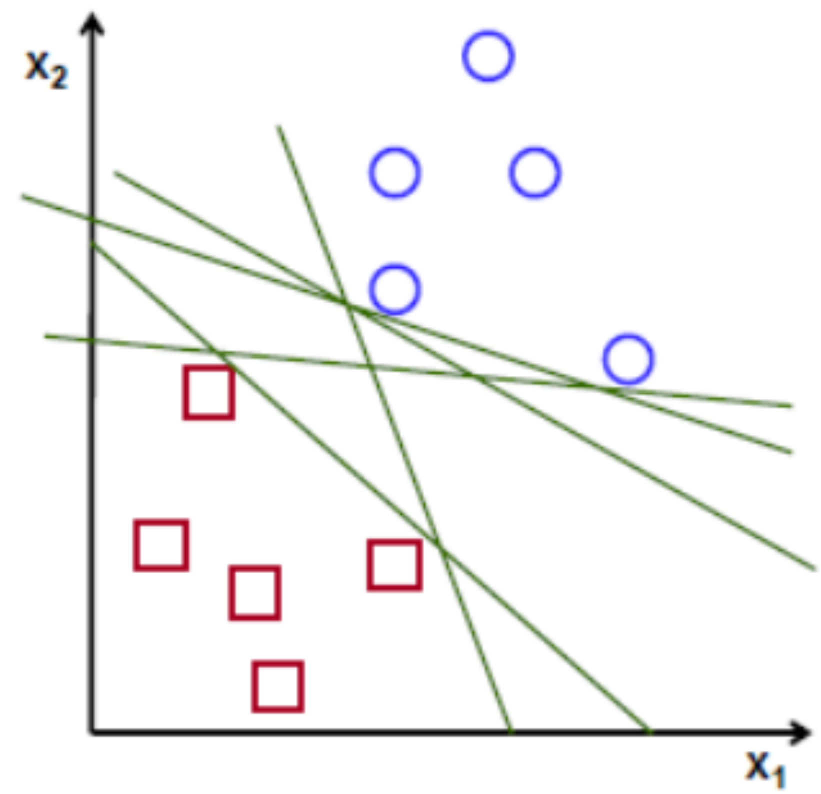
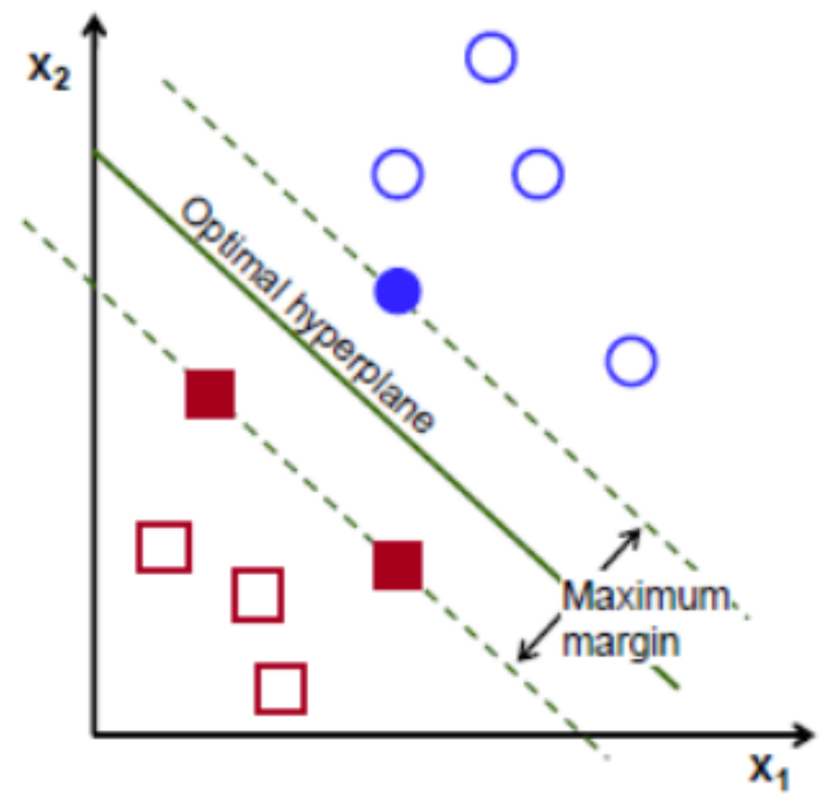
Support Vector Machine (SVM)

Group 23

Run Zeng 20794087

Yewei Li 20759952

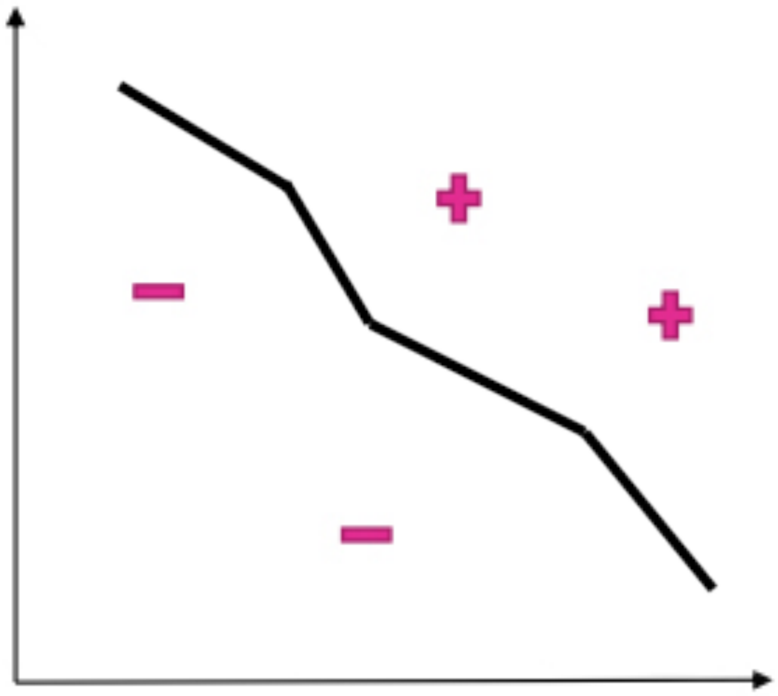




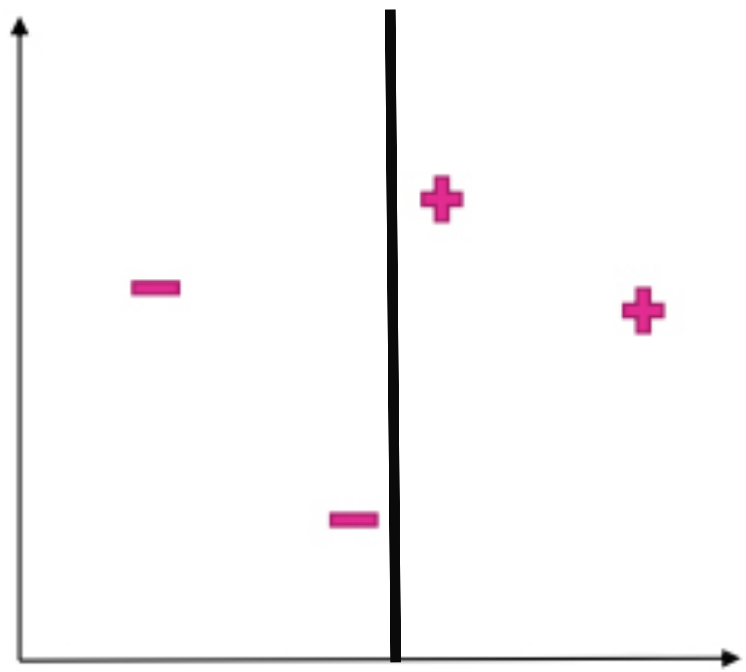
UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

Why we need SVM

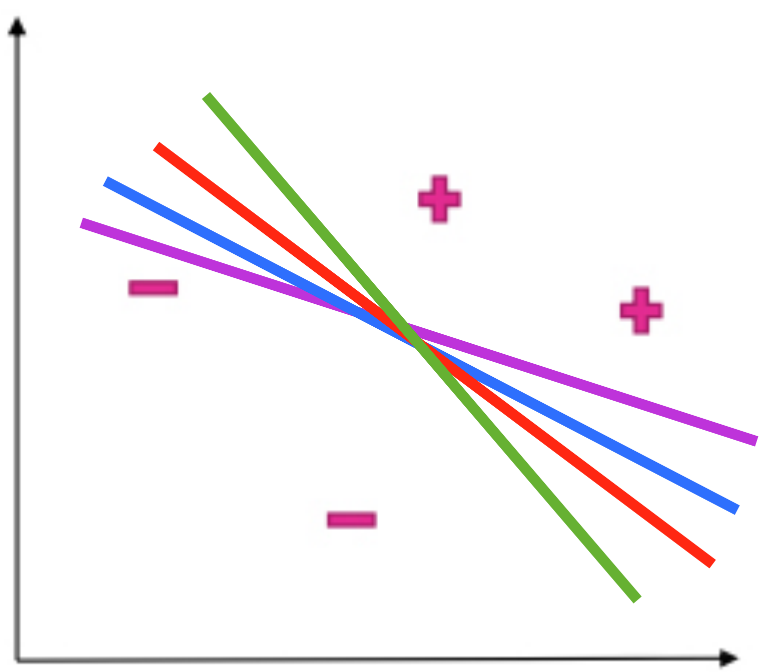
kNN when k=1



Decision tree



Which is the best fit line?

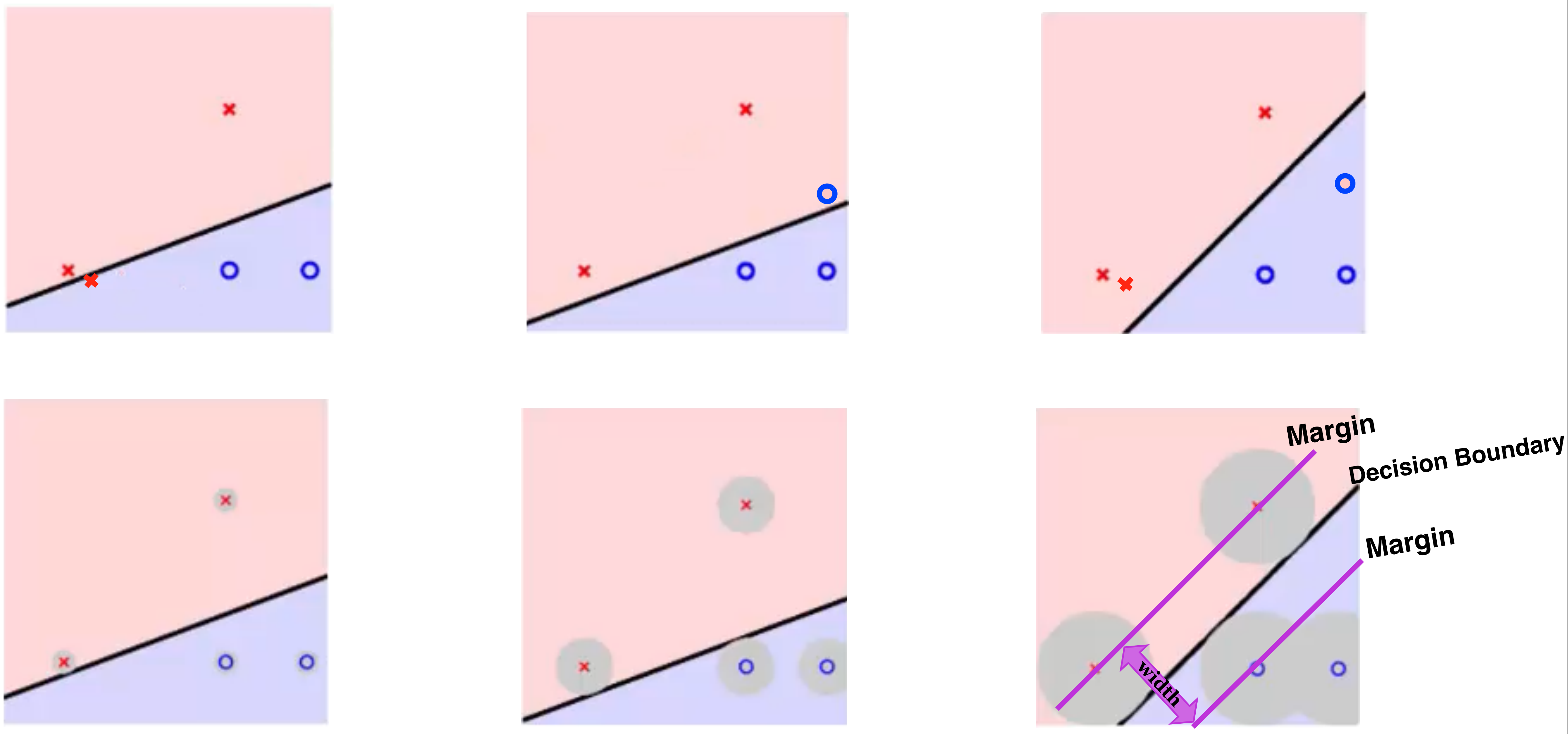


UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

Introduction of SVM

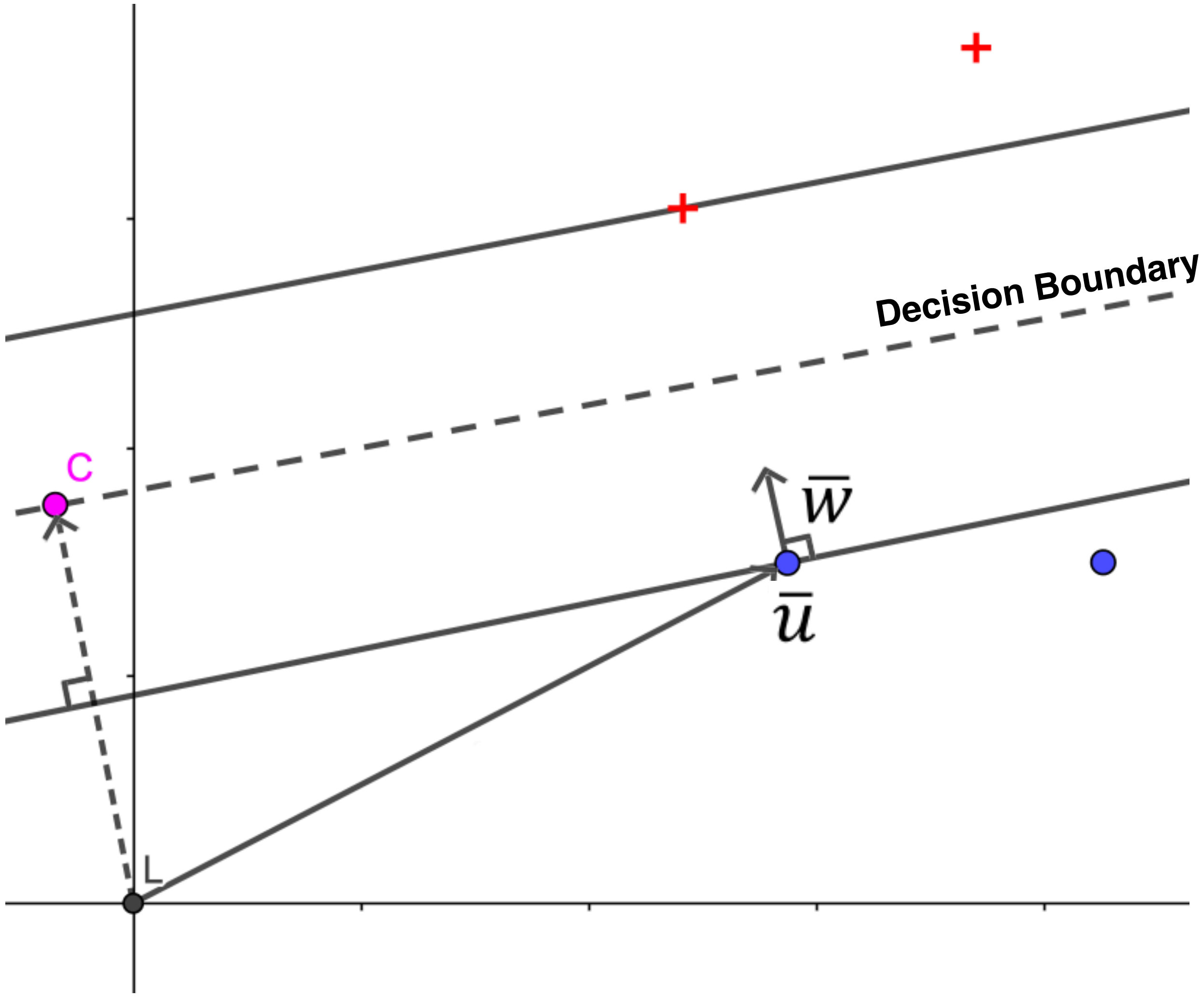
How to find the best decision boundary ?

The goal of SVM



- The goal for SVM is to find a decision boundary which maximizes the distance between decision boundary and the closest data to it.

Optimal Decision Boundary



- $\overline{w}$ is a unit vector perpendicular to decision boundary.

- The sample $\overline{u}$ is positive when $\overline{w}.\overline{u} \geq c$
- This equals $\overline{w}.\overline{u} + \mathrm{b} \geq 0$

Optimal Decision Boundary

- A better restriction

$\bar{w}.\overline{x_+} + \mathrm{b} \geq 1$ then the sample is positive

$\bar{w}.\overline{x_-} + \mathrm{b} \leq -1$ then the sample is negative

To combine the two equations:

y_i s.t. $y_i = +1$ for positive samples

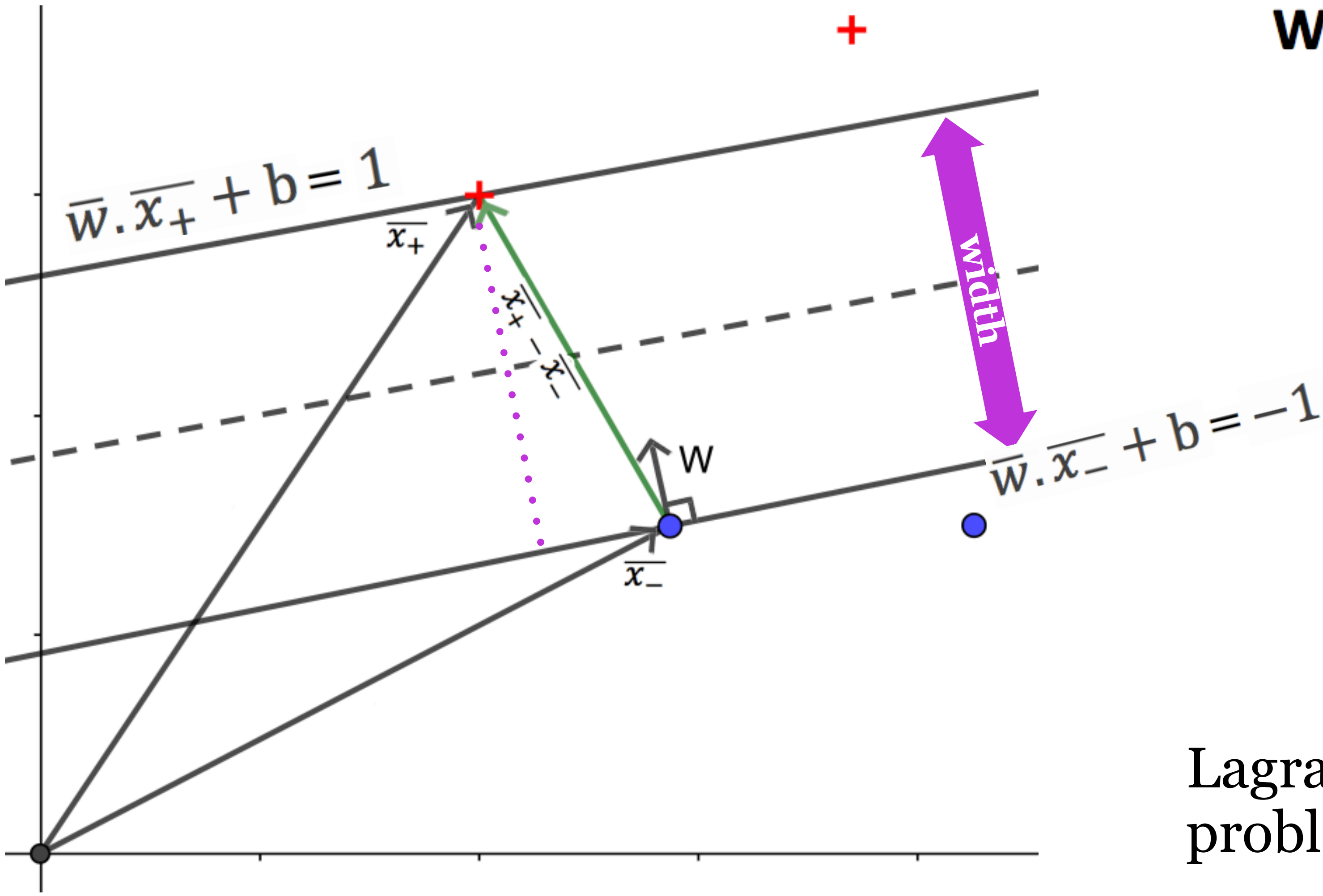
$y_i = -1$ for negative samples

$$y_i(\bar{w}.\overline{x_+} + \mathrm{b}) \geq 1$$
$$y_i(\bar{w}.\overline{x_-} + \mathrm{b}) \geq 1$$

$$\rightarrow$$

$$y_i(\bar{w}.\bar{x_i} + \mathrm{b}) - 1 \geq 0$$

SVM



$$\text{Width} = \frac{w}{\|w\|} \cdot (\overline{x_+} - \overline{x_-})$$

$$\text{Width} = \frac{1\text{-b-(-1-b)}}{\|w\|} = \frac{2}{\|w\|}$$

$$\text{Max}\ \frac{2}{\|w\|} \rightarrow \text{Max}\ \frac{1}{\|w\|}$$

$$\rightarrow \text{Min}\ \frac{1}{2} \| w \|^2 \checkmark$$

$$y_i(\overline{w}.\overline{x_i} + \text{b}) - 1 = 0 \checkmark$$

Lagrange multiplier to solve this optimization problem

Optimization

$$\boxed{\mathbf{J(w,b) = L = \frac{1}{2} \parallel w \parallel^2 - \sum \alpha^i \left[y_i(\bar{w}.\bar{x}_i + b) - 1\right]}}$$

$$\frac{\partial L}{\partial \bar{w}} = \bar{w} - \sum \alpha^i y_i x_i = 0 \longrightarrow \bar{w} = \sum \alpha^i y_i x_i$$

$$\frac{\partial L}{\partial b} = -\sum \alpha^i y_i = 0 \longrightarrow \sum \alpha^i y_i = 0$$

$$\boxed{L = \sum \alpha^i - \frac{1}{2}\left(\sum\sum \alpha^i \alpha^j y_i y_j x_i \cdot x_j\right)}$$

- We are able to solve the equation with numerical analytic method, so that we get all the α_i α_j
- So for the new unknown sample $\bar{u}$

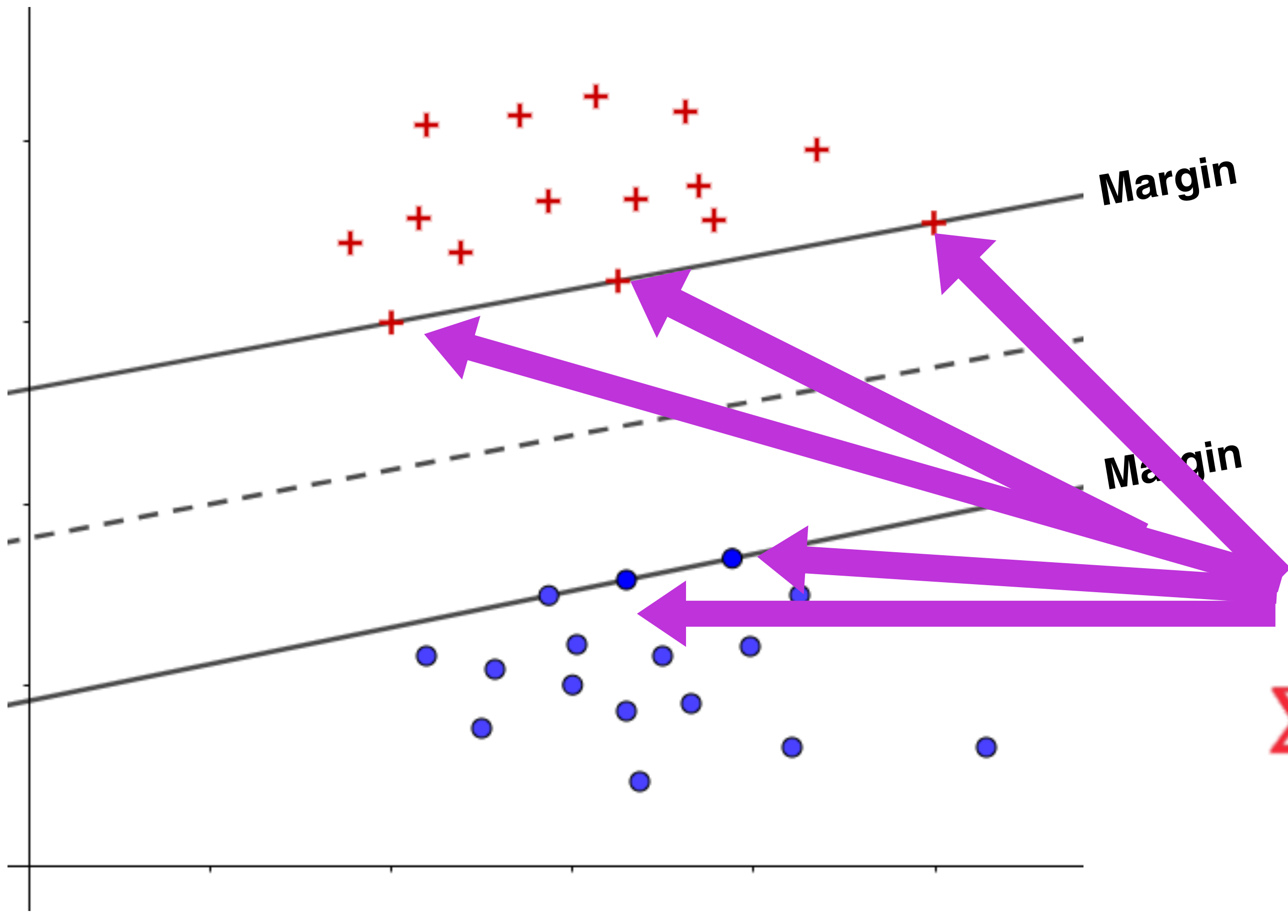
$$\bar{w}.\bar{u} + \mathrm{b}$$

$\sum \alpha^i y_i \bar{x}_i \cdot \bar{u} + b \geq 0$ Then class label is positive

≤ 0 class label is negative

UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

Support Vectors



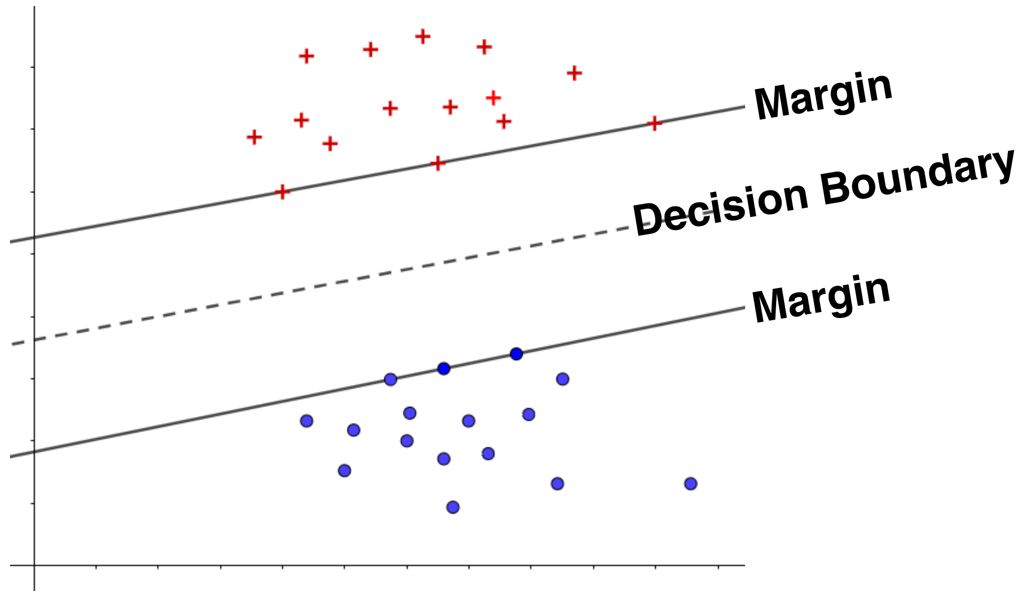
$\alpha^i > 0$ For support vector samples

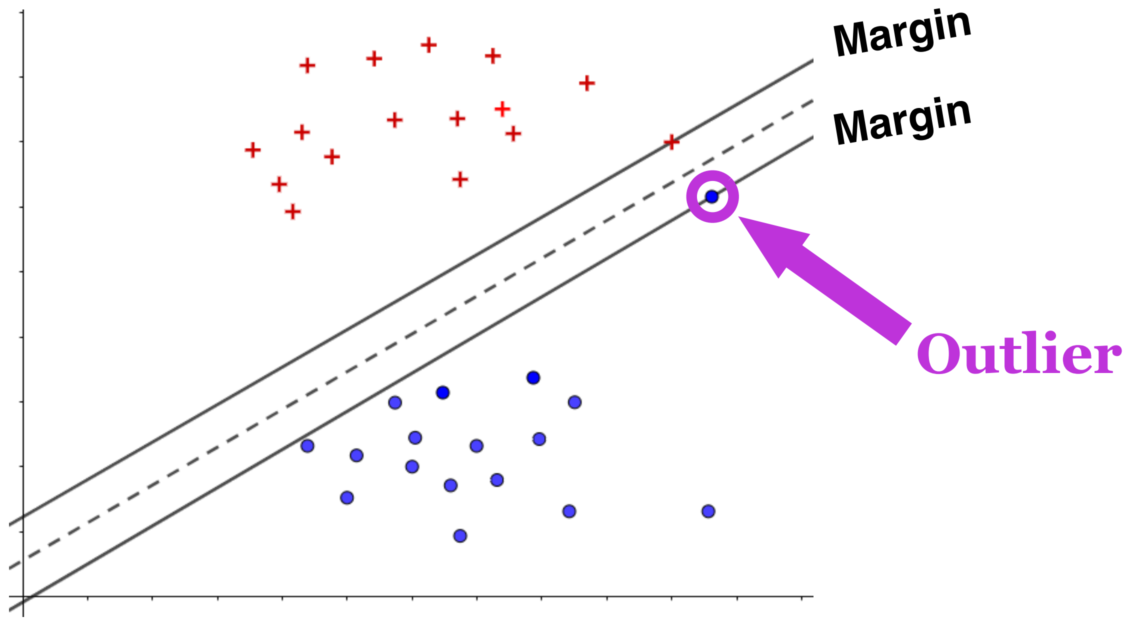
$\alpha^i = 0$ For other samples

Support Vectors

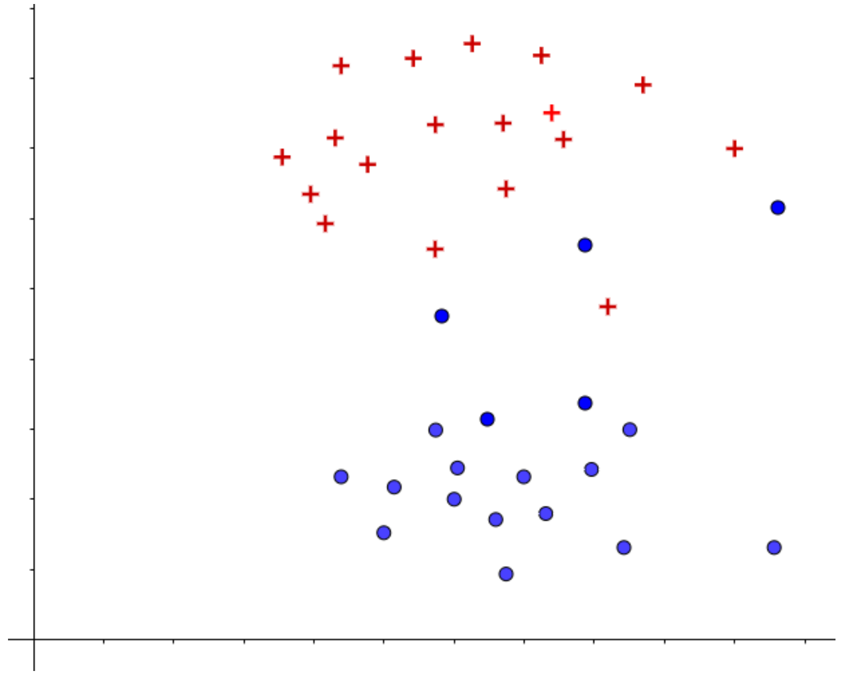
$$\sum \alpha^i y_i \bar{x}_i \cdot \bar{u} + b \geq 0 \quad \text{Then class label is positive}$$
$$\leq 0 \quad \text{class label is negative}$$

Hard-Margin SVM





An example could not be linearly separable



UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

Soft-Margin SVM

- Recall one constraints we draw to deal with the Hard-margin SVM problem
  $y^{(i)}(w^T x^{(i)} + b) \geq 1$
- Soft-Margin SVM introduces slack variables ξ_i that relax the constraint
  $y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi_i \qquad \forall i = 1 \ldots N$
- Then the previous optimization problem is defined as:

$$\text{Minimize} \quad J(w) = \frac{1}{2}||w||^2 + \boldsymbol{C}\sum_{i=1}^{N}\xi_i$$

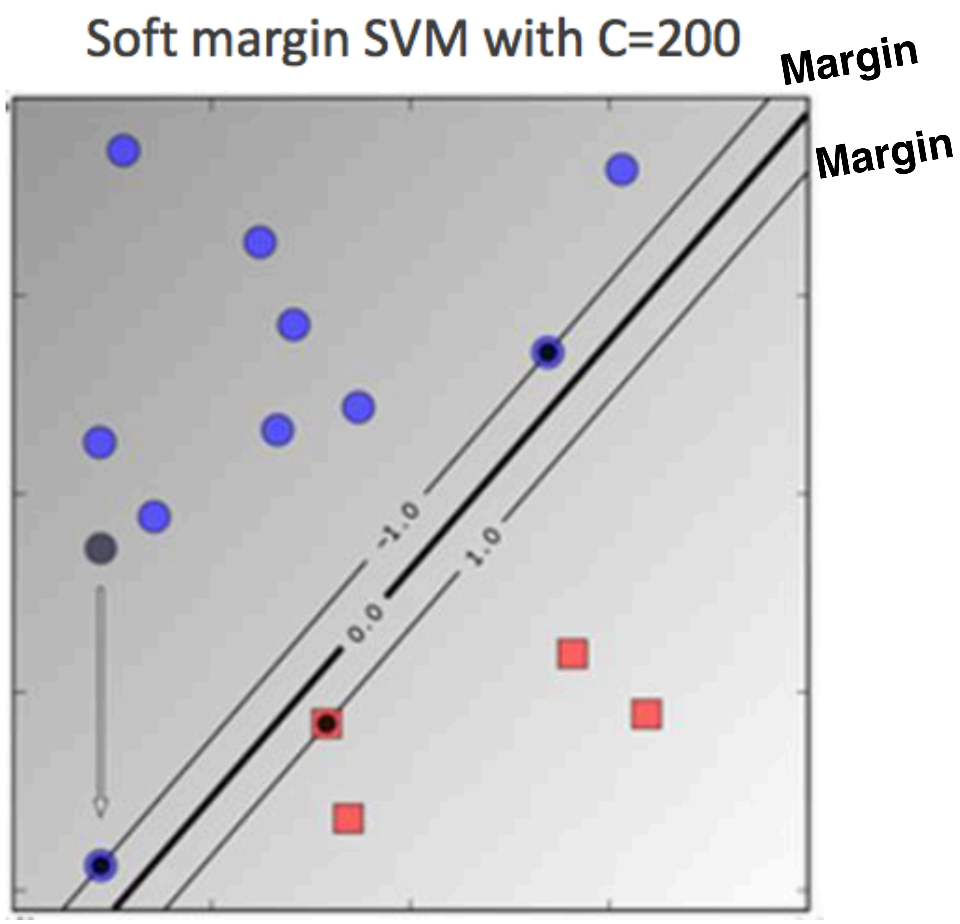
$$\text{Subject to} \quad \begin{cases} y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi_i & \forall i \\ \xi_i \geq 0 & \forall i \end{cases}$$

- Parameter C:
  It controls the trade-off between complexity and miss-classification

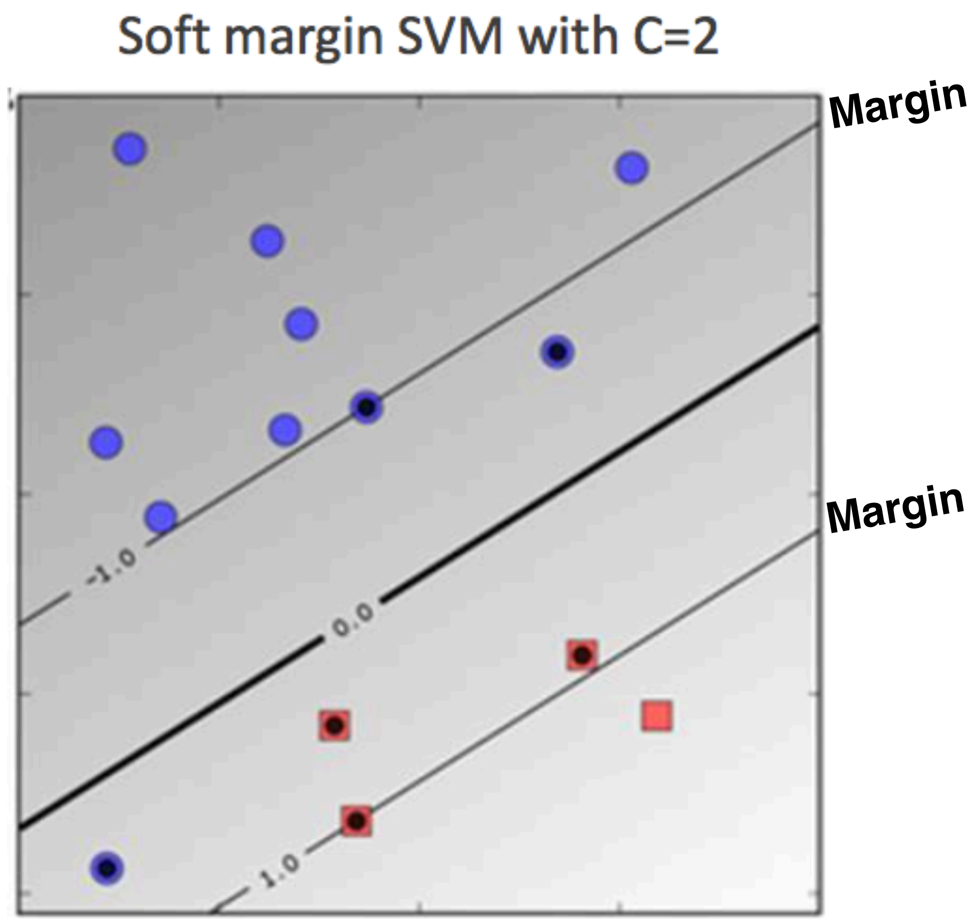
UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

Effect of Penalization Factor C

Larger C: solutions with lower miss-classification error



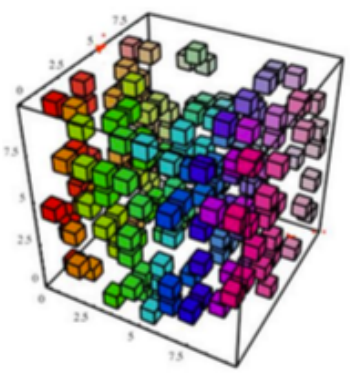
Smaller C: solutions with some miss-classification



[Figure by Ben-Hur, et. al. 2008]

Non-linear SVM

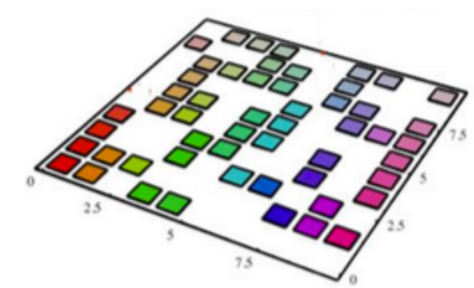
3D



$\boldsymbol{x}$: inpute samples in a 3D feature space

Feature extraction

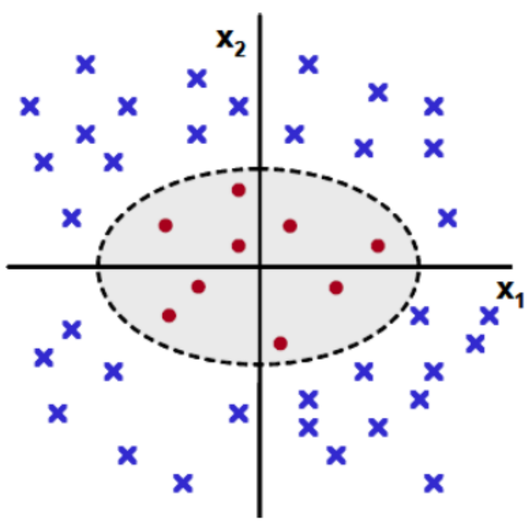
$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & & w_{1D} \\ w_{21} & w_{22} & \cdots & & w_{1D} \\ \vdots & \vdots & & \ddots & \vdots \\ w_{d1} & w_{d2} & \cdots & \cdots & w_{dD} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \\ x_D \end{bmatrix}$$



2D

$\boldsymbol{y}$: transformed Samples in a 2D feature space

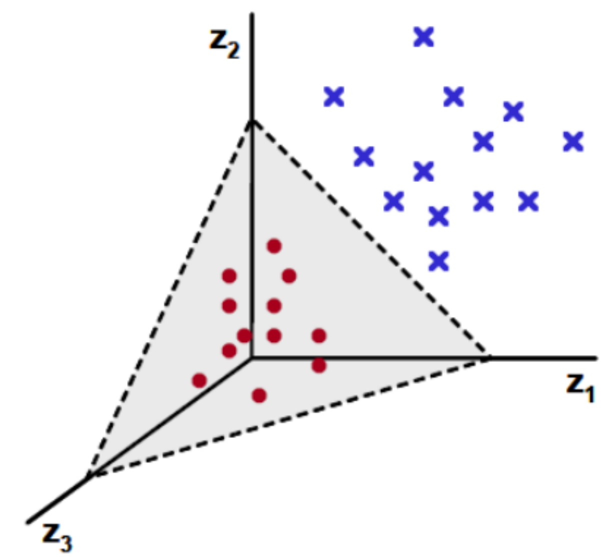
2D



Non-linear function

$$\varphi: \; R^2 \rightarrow R^3$$

$$(x_1, x_2) \;\mapsto (z_1, z_2, z_3) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$



3D

Kernel SVM = Non-linear SVM

- Decision Function of Non-linear SVM will be

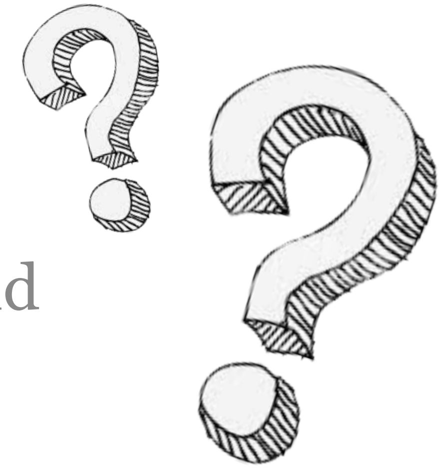
$$h(x) = sign(w^T \varphi(x) + b) = sign\left(\sum_{i=1}^{m} \alpha_i y^{(i)} \varphi\left(x^{(i)}\right)^T \varphi(x) + b\right)$$

- Kernel Function

$$K\left(x^{(i)}, x^{(j)}\right) = \varphi(x^{(i)})^T . \varphi(x^{(j)})$$

$$h(x) = sign\left(\sum_{i=1}^{m} \alpha_i y^{(i)} K(x^{(i)}, x) + b\right)$$

Extensions of SVM



what issues when apply SVM to real world

Advantages

- Can solve linear non-separate problem by mapping datas to high dimensions via kernel functions

- Deleting and adding the datas which are not support vectors do not affect modeling

Disadvantages

- Very difficult to choose Kernel Functions and parameter (c, g)

- Perform poorly in imbalanced samples

UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

Development

- Deal with unsupervised learning based on SVM

Example: Marketing segmentation of a drink company

Support vectors clustering(unsupervised learning)

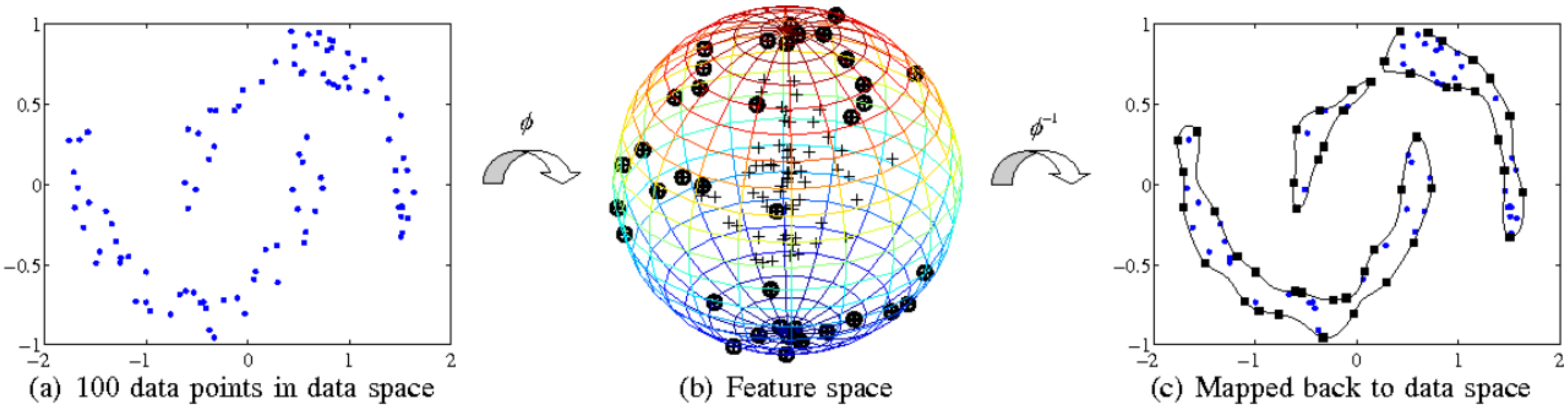


Figure 1. (a) Original data (b) Map to high dimension and minimal enclosing sphere (c) Map back to achieve clustering

Development

- Deal with semi-supervised learning based on SVM

Example: Hand written recognition

Figure 2. Hand-written digits data sets(0, 6, 8, 9)

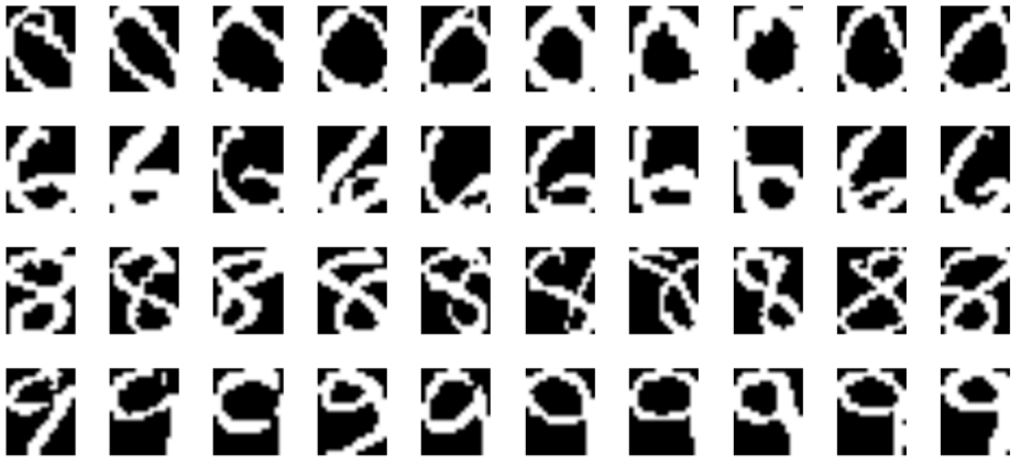


Figure 3. Unlabelled hand written digits data sets

Transductive support-vector machines(semi-supervised learning)

Development

- Deal with multi-classification problem based on SVM

Classes of fruits' images

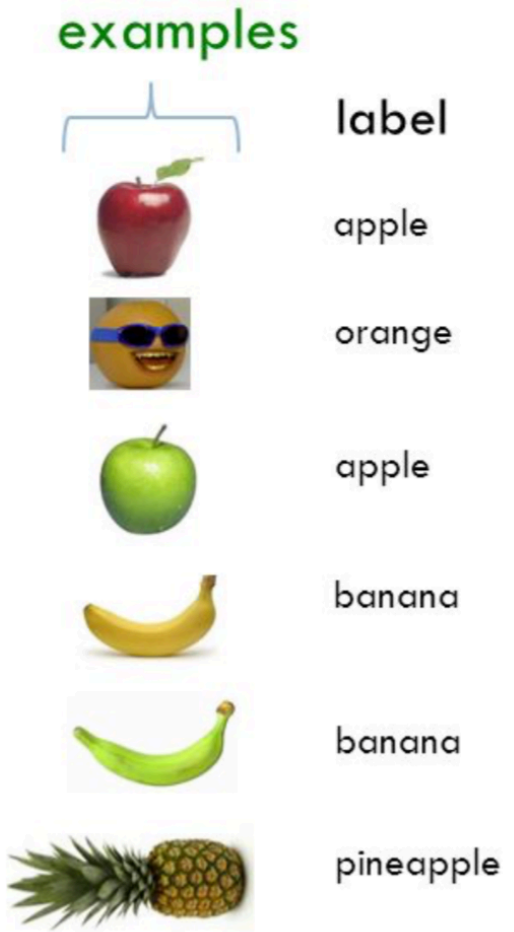


Figure 4. multi-classification problem of fruits

Multi-class SVM

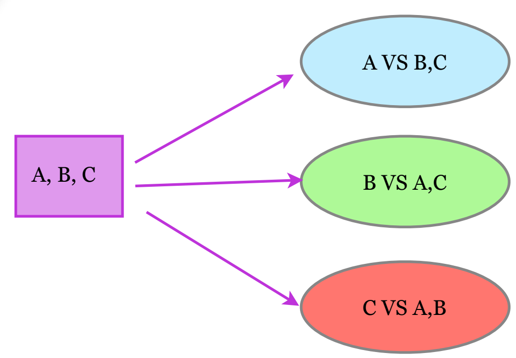


Figure 5. **one-versus-rest(OVR)** for a problem with three classes

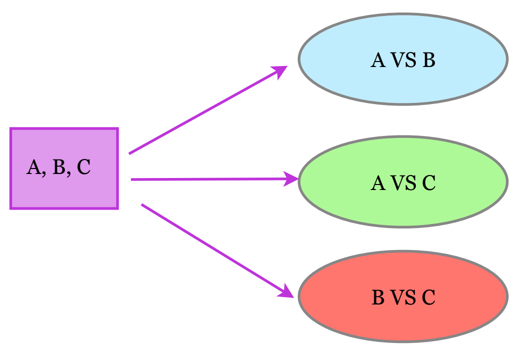


Figure 6. **one-versus-one(OVO)** for a problem with three classes

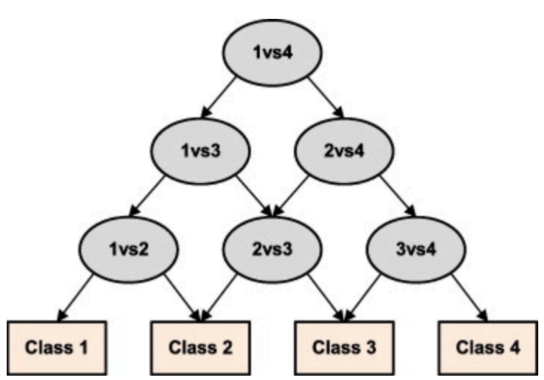


Figure 7. **DAGSVM** with four different classes

UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

Reference

- Support-vector machine Wikipedia
- Asa Ben-Hur, David Horn, Hava T. Siegelmann, Vladimir Vapnik, Support Vector Clustering, Journal of Machine Learning Research 2, pp.125-137, 2001.
- Jih-Jeng Huang, Gwo-Hshiung Tzeng, Chorng-Shyong Ong, Marketing segmentation using support vector clustering, Expert Systems with Applications 32, pp.313–317, 2007.
- Linli Xu, Dale Schuurmans, Unsupervised and Semi-supervised Multi-class Support Vector Machines, 2005.
- Chen YS, Wang GP, Dong SH. A progressive transductive inference algorithm based on support vector machine. Journal of Software, 14(3):451~460, 2003.
- Alexander Statnikov, Constantin F. Aliferis, Ioannis Tsamardinos, Douglas Hardin, Shawn Levy, A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosisBioinformatics, Volume 21, Issue 5, pp.631–643, 2005.
- Ben Aisen, A Comparison of Multiclass SVM Methods, 2006.
- Andraw, https://liuhongjiang.github.io/hexotech/2012/12/29/svm-ocr/, Hand Written Pattern Recognization base on SVM, 2012.
- Christophere J.C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, 2, pp.121–167, 1998.
- Mohammad Javad Shafiee, Support Vector Machine (SVM) slides, winter 2019.

Q & A :

UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

Group 23

thank you !

Disadvantages

- Perform poorly in unbalanced samples

Solution:

Example

Disadvantages

- Very difficult to choose Kernel Functions and parameter (c, g)

Linear kernel: $K(x, x') = \left(\mathrm{x}^{\mathrm{T}} x'\right)^{2}$

Polynomial kernel: $K(x, x') = \left(\mathrm{x}^{\mathrm{T}} x' + 1\right)^{\mathrm{d}}$

Radial basis function kernel: $K(x, x') = \exp\left(-\frac{||x - x'||^{2}}{2\sigma^{2}}\right)$

Sigmoid kernel: $k(\vec{u}, \vec{v}) = \tanh(\gamma \vec{u}^{T} \cdot \vec{v} + coef0)$

$c \to \infty$: model $\to$ hard-margin
$c \to 0$: model $\to$ invalid

For same samples, different kernel functions perform similarly.
(c,g) depends on cross-validation

Why we need SVM?

- Lorem ipsum dolor sit amet, consectetuer adipiscing elit
- Maecenas porttitor congue massa
- Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna
  eros quis urna
- Nunc viverra imperdiet enim.

PAGE TITLE HERE

- Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.
- Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.
- Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede. Mauris et orci.

PAGE TITLE HERE

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus non sem semper, rhoncus quam nec, suscipit erat. In non scelerisque est. Cras eu sodales mauris. Integer a magna non tellus vestibulum venenatis sed vel urna. Morbi scelerisque et ligula sed maximus. Sed tincidunt lacus id libero hendrerit condimentum. Praesent euismod est finibus metus euismod mattis. Aliquam malesuada interdum nisl sed tincidunt. Sed et eleifend urna. Curabitur tempor vulputate leo vitae consequat.

TWO COLUMN CONTENT SLIDE LAYOUT

- Lorem ipsum dolor sit amet, consectetuer adipiscing elit
- Maecenas porttitor congue massa
- Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna
- Nunc viverra imperdiet enim.

- Lorem ipsum dolor sit amet, consectetuer adipiscing elit
- Maecenas porttitor congue massa
- Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna
- Nunc viverra imperdiet enim.

$$\bar{w}.\bar{u} + \mathrm{b} \geq 0$$

$$\sum \boldsymbol{\alpha^{i}}\, \boldsymbol{y_i} \boldsymbol{\bar{x}_i} \cdot \bar{u} + \boldsymbol{b}$$

TWO COLUMN COMPARISON CONTENT SLIDE LAYOUT

Subheading left side

- Lorem ipsum dolor sit amet, consectetuer adipiscing elit
- Maecenas porttitor congue massa
- Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna
- Nunc viverra imperdiet enim.

Subheading right side

- Lorem ipsum dolor sit amet, consectetuer adipiscing elit
- Maecenas porttitor congue massa
- Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna
- Nunc viverra imperdiet enim.

PAGE TITLE HERE

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna. Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede. Mauris et orci. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

IMAGE CREDIT: NAME HERE

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus non sem semper, rhoncus quam nec.

PAGE TITLE HERE

EXAMPLE.1

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus non sem semper, rhoncus quam nec.

EXAMPLE.2

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus non sem semper, rhoncus quam nec.

EXAMPLE.3

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus non sem semper, rhoncus quam nec.

CONTEXT OR THEME

“Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus non sem semper, rhoncus quam nec, suscipit erat. In non scelerisque est. Cras eu sodales mauris.”

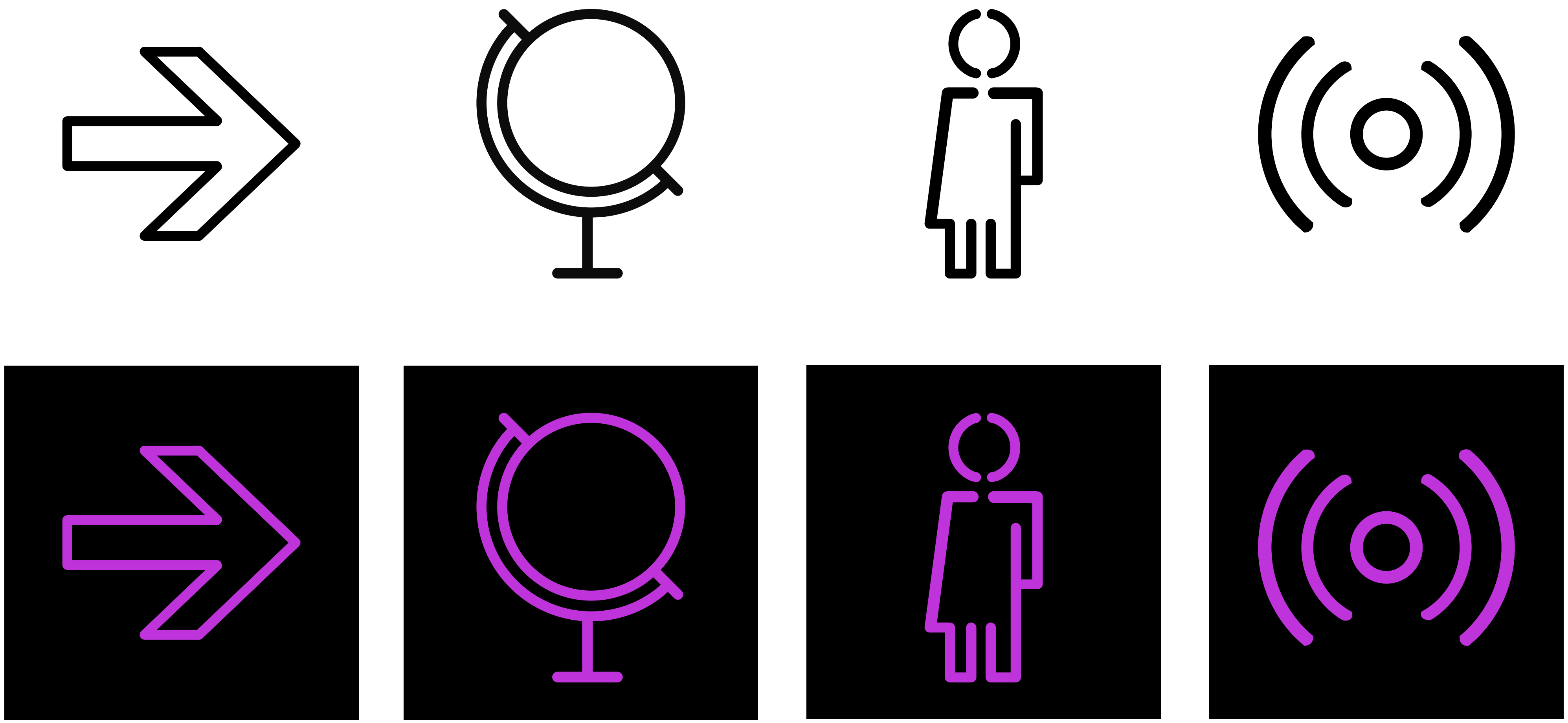
NAME HERE

CONTEXT OR THEME

“Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus non sem semper, rhoncus quam nec, suscipit erat. In non scelerisque est. Cras eu sodales mauris.”

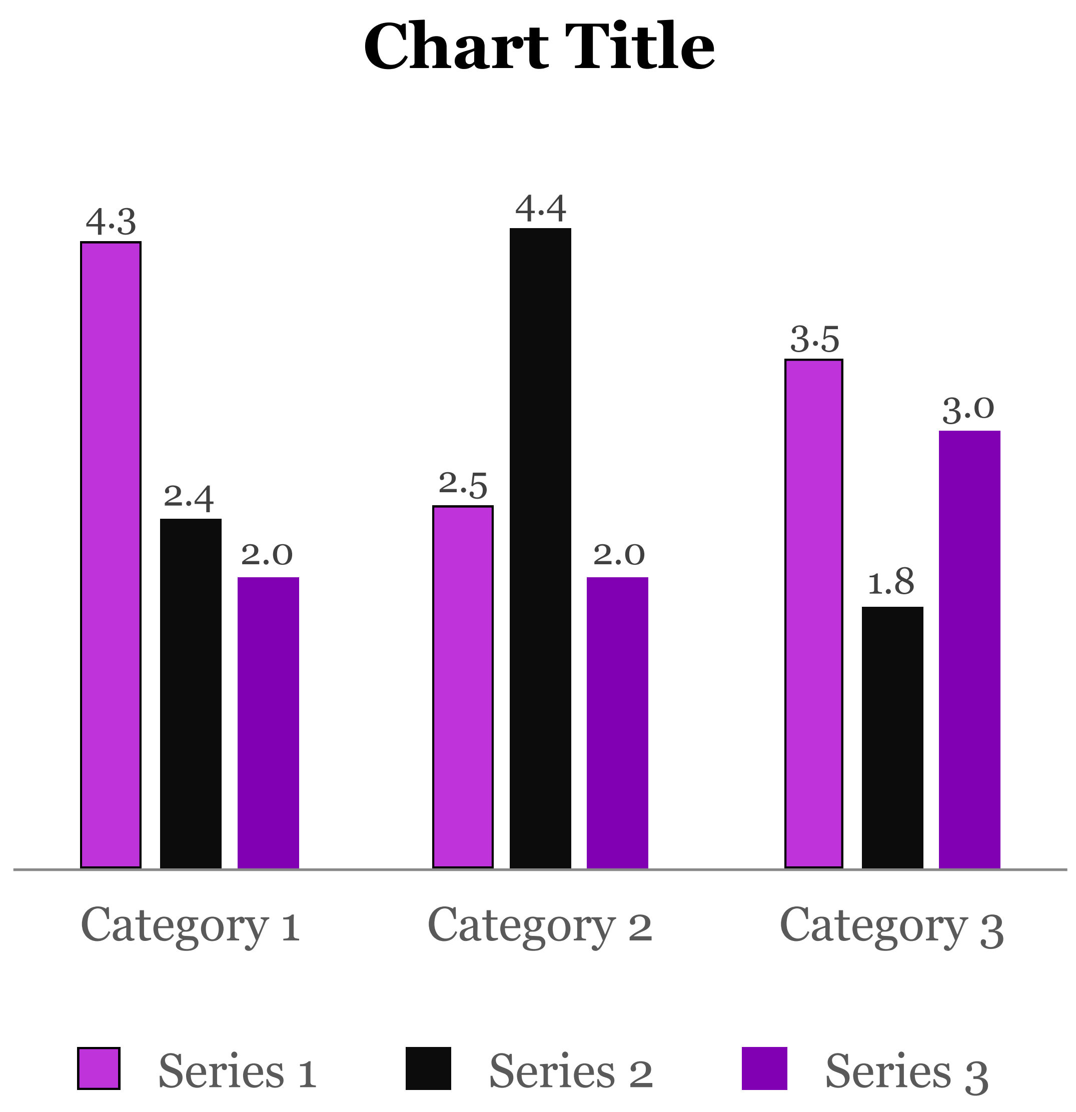
NAME HERE

INFOGRAPHIC SYMBOLS / STYLE



UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

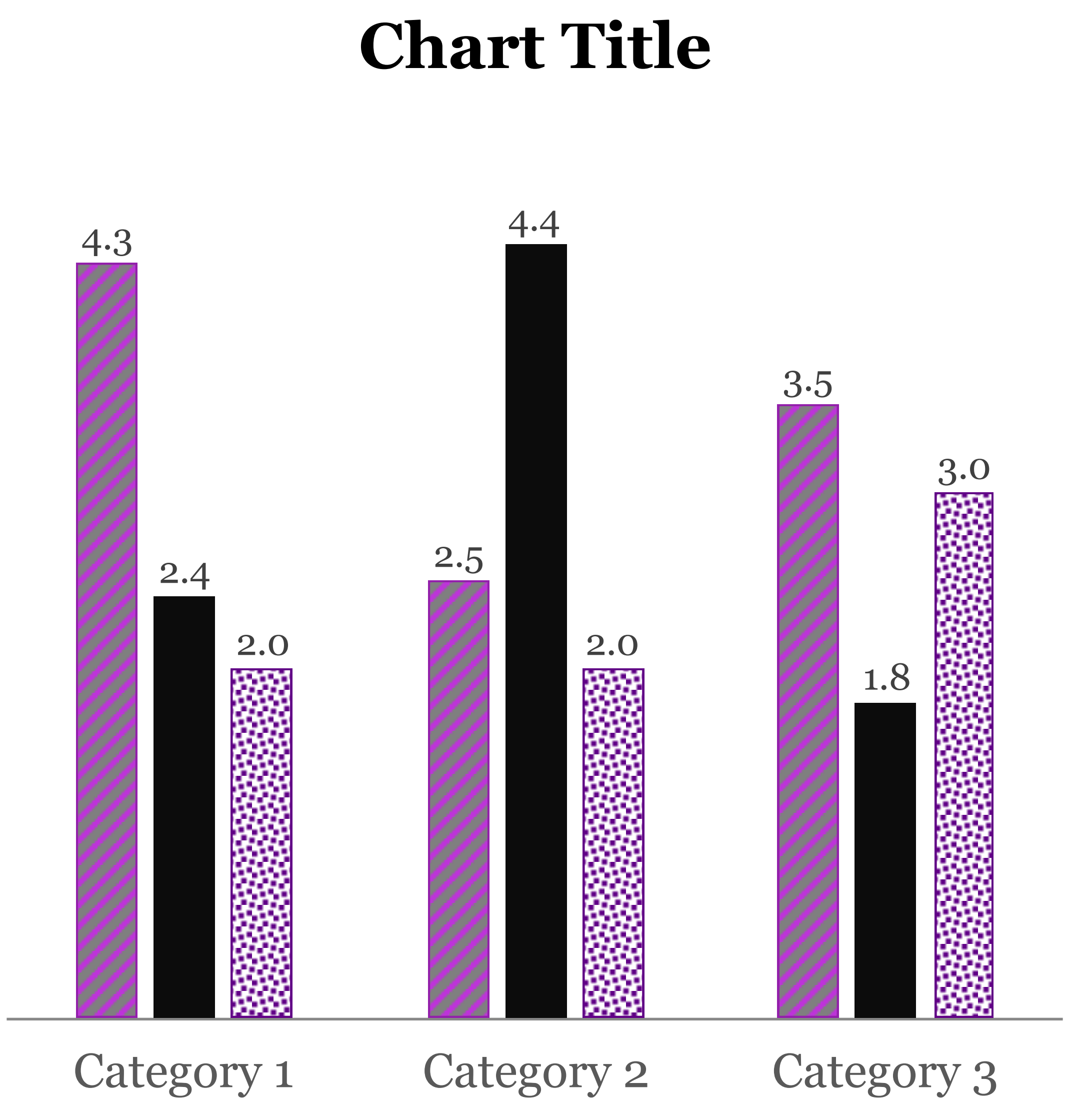
MASTER SLIDE CHART OPTION 1



- Lorem ipsum dolor
  sit amet, consectetuer adipiscing elit.
  Maecenas porttitor
  congue massa.

- Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo.

MASTER SLIDE CHART OPTION 1 (accessible version)

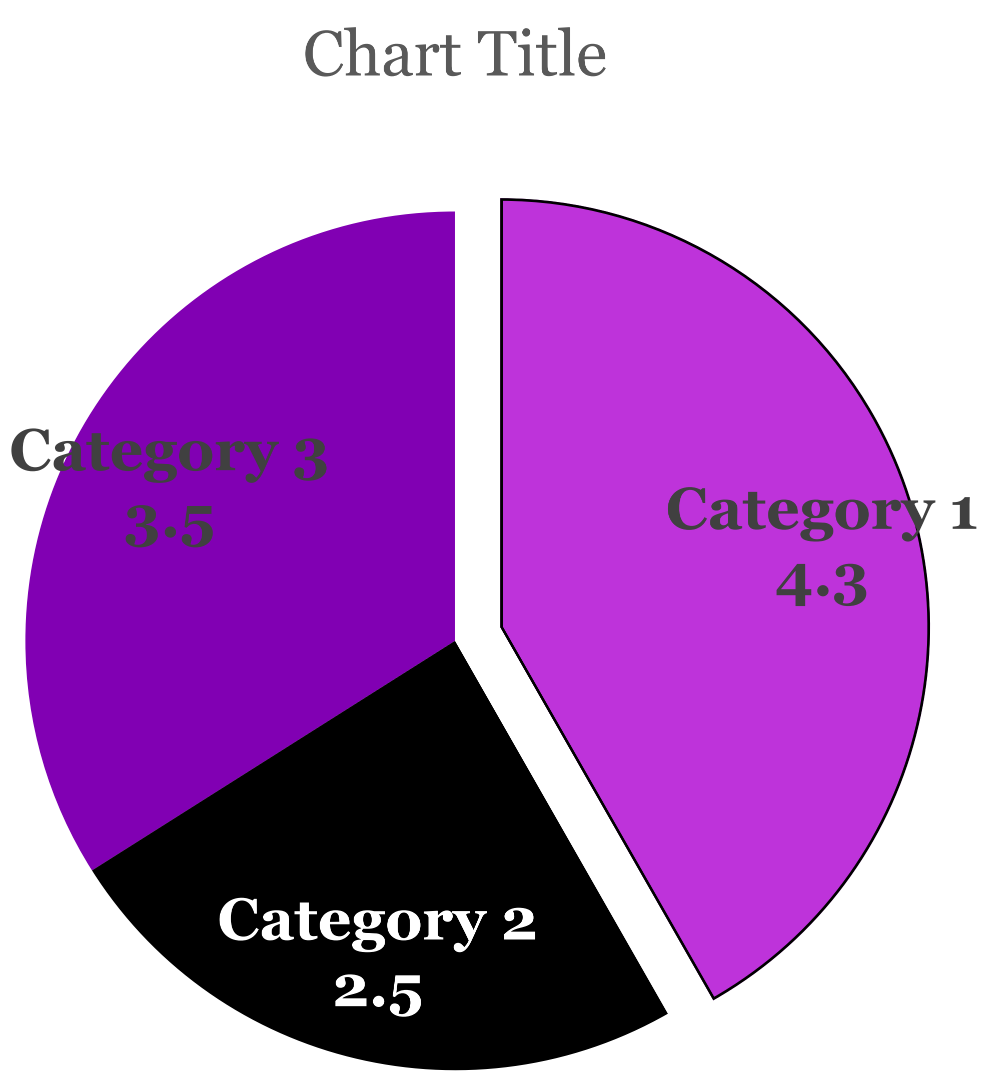


- Lorem ipsum dolor
  sit amet, consectetuer adipiscing elit.
  Maecenas porttitor
  congue massa.

- Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo.

UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

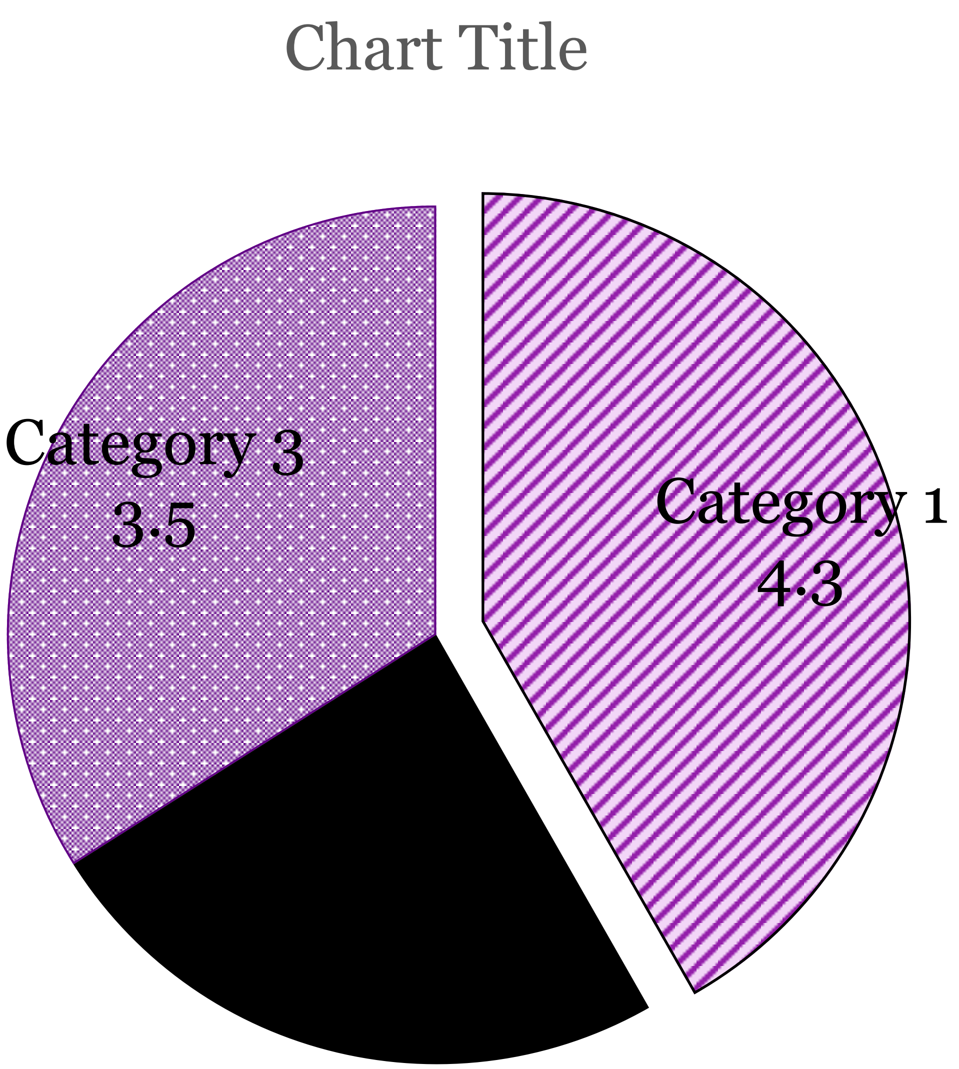
MASTER SLIDE CHART OPTION 2



- Lorem ipsum dolor
  sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa.
- Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo.

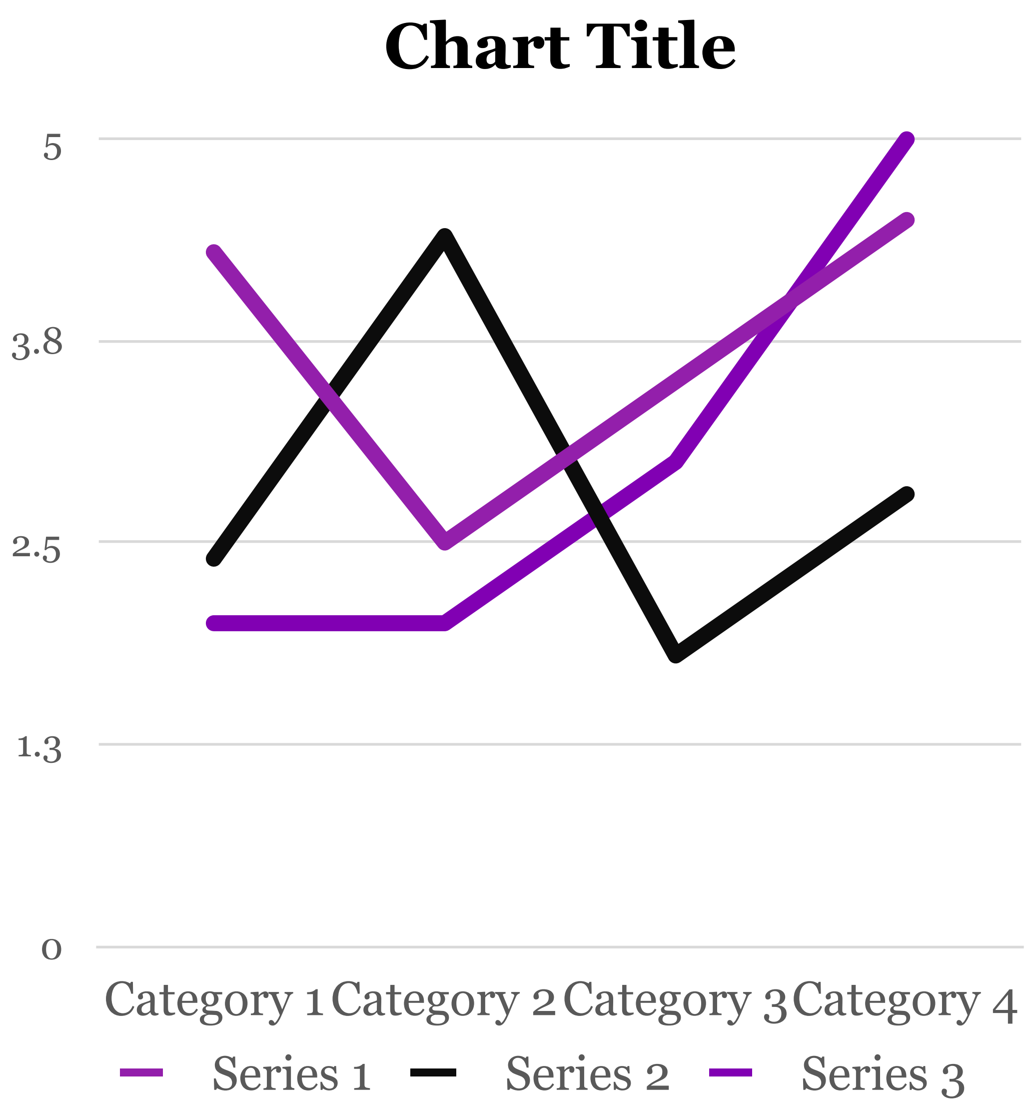
UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

MASTER SLIDE CHART OPTION 2 (accessible version)



- Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa.
- Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo.

MASTER SLIDE CHART OPTION 3

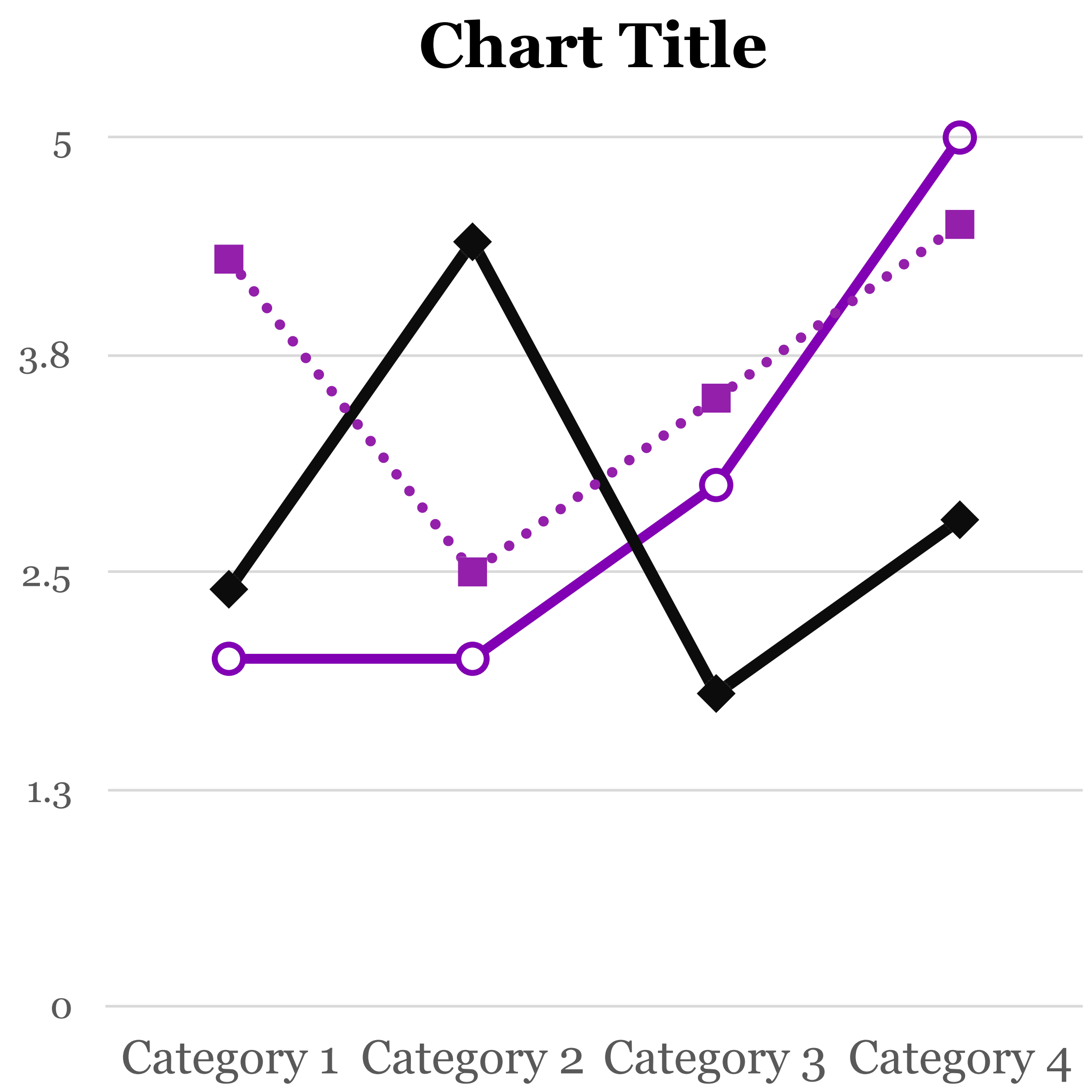


- Lorem ipsum dolor
  sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa.

- Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo.

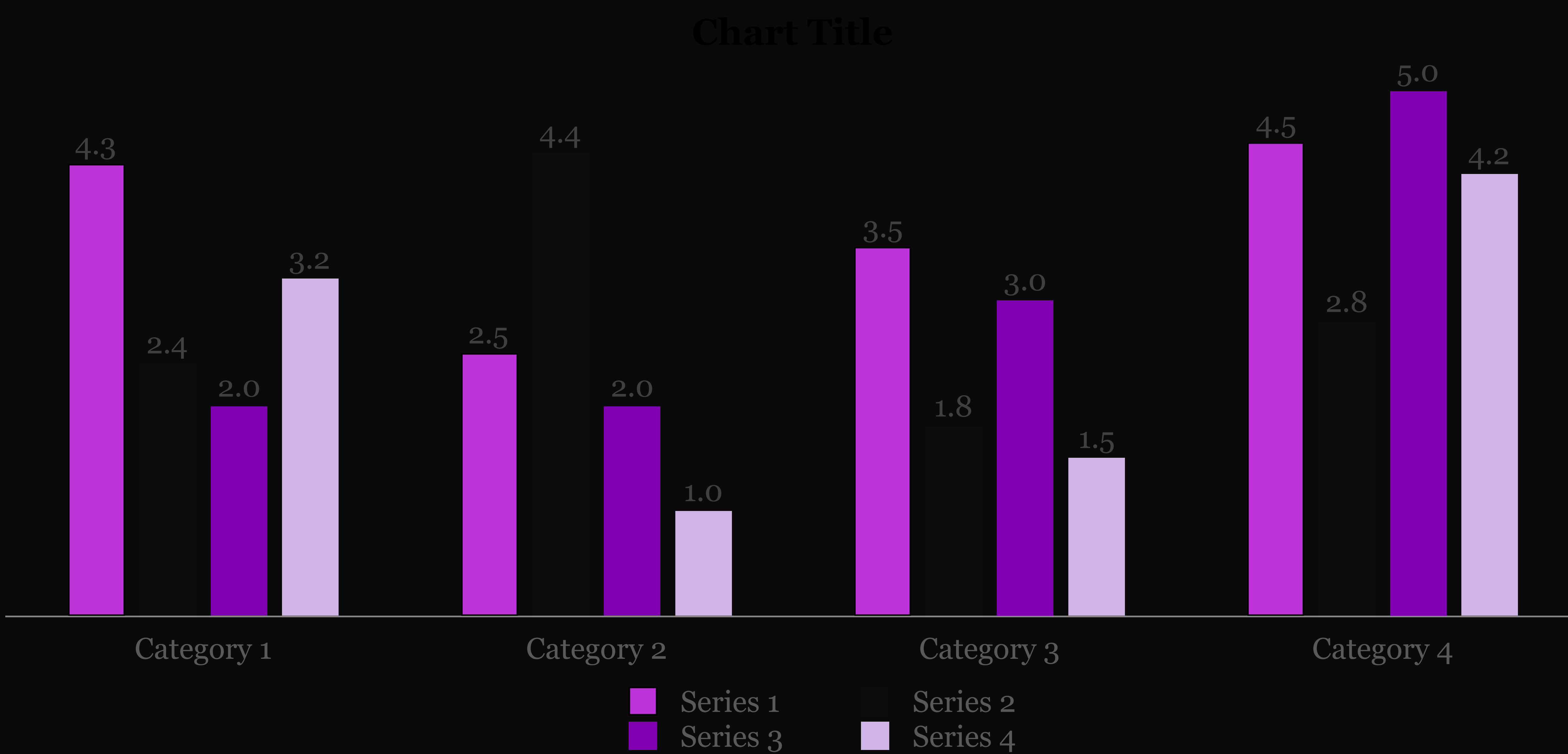
UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

MASTER SLIDE CHART OPTION 3 (accessible version)



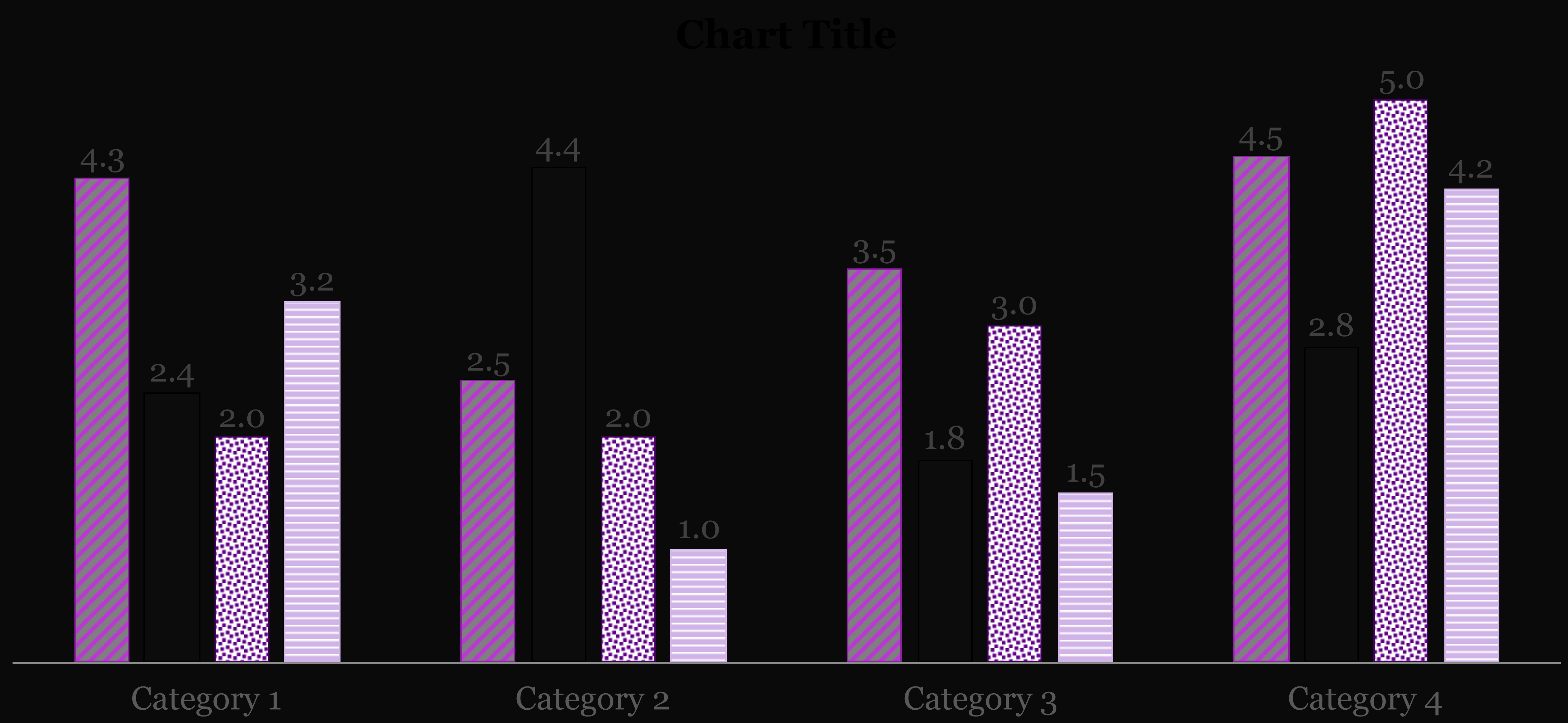
- Lorem ipsum dolor
  sit amet, consectetuer adipiscing elit.
  Maecenas porttitor congue massa.
- Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo.

MASTER SLIDE CHART OPTION 4



UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

MASTER SLIDE CHART OPTION 4 (accessible version)



UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

MENU SYSTEM EXAMPLE

MENU SYSTEM EXAMPLE

MENU SYSTEM EXAMPLE

UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING


UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING


UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING


UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING


UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING


UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING