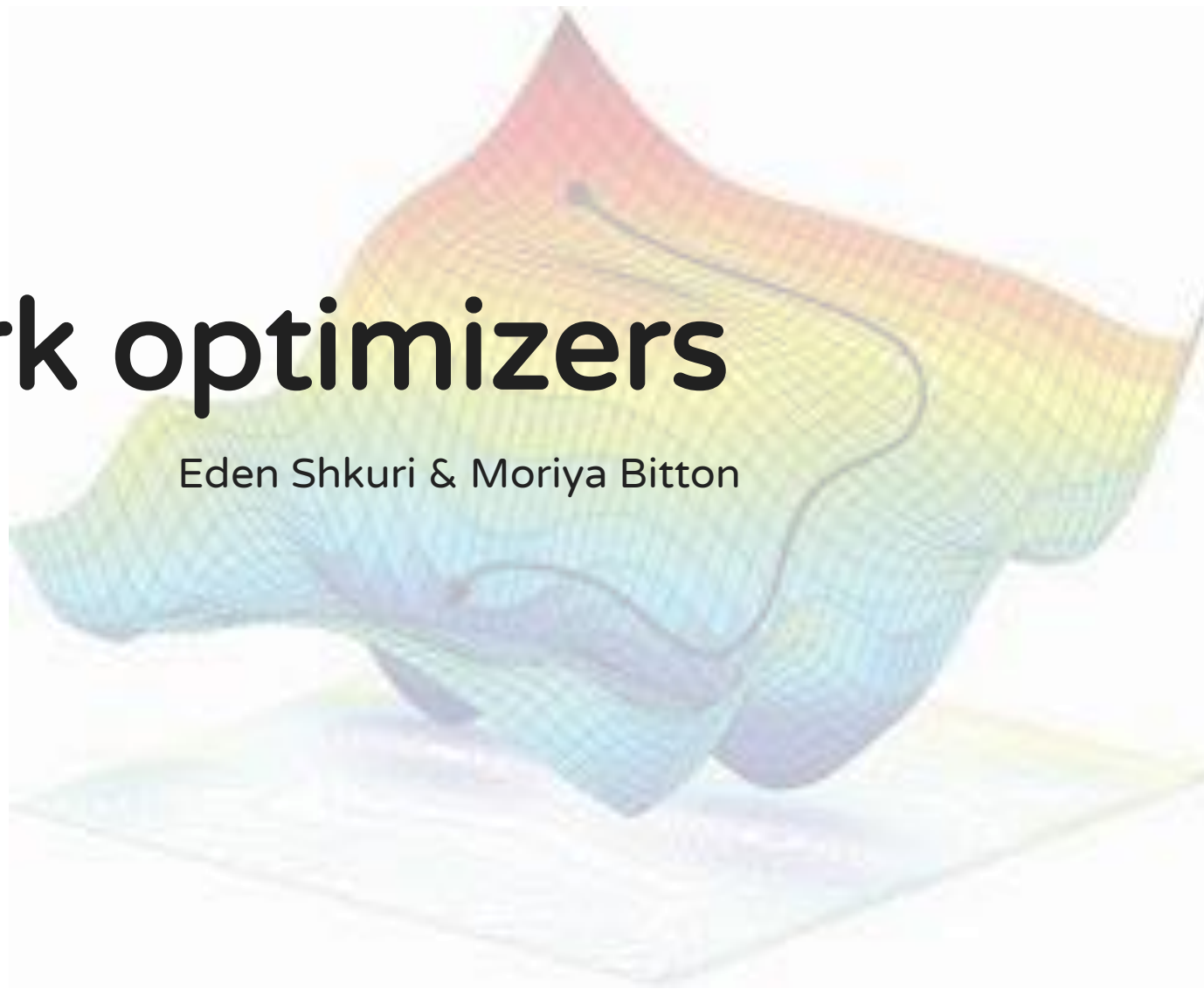
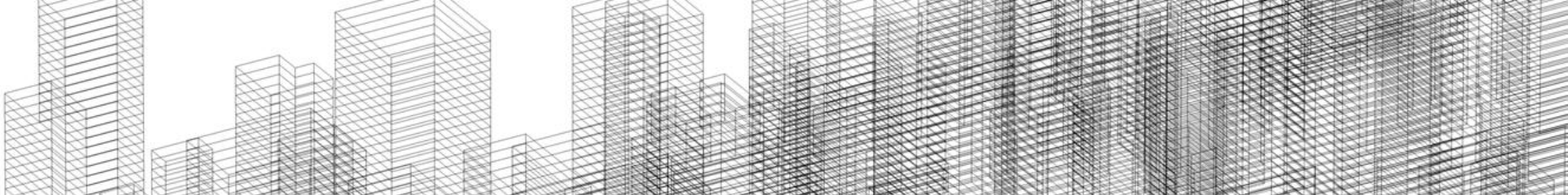


Network optimizers

Eden Shkuri & Moriya Bitton





what is an optimizer?

אופטימיזר הוא פונקציה או אלגוריתם העוזר להפחית את ה-Loss ולשפר את הדיוק בהגעה לנקודת המינימום ע"י שינוי התכונות של רשת הנוירונים, כגון משקלים וקצב למידה.

Our main goal

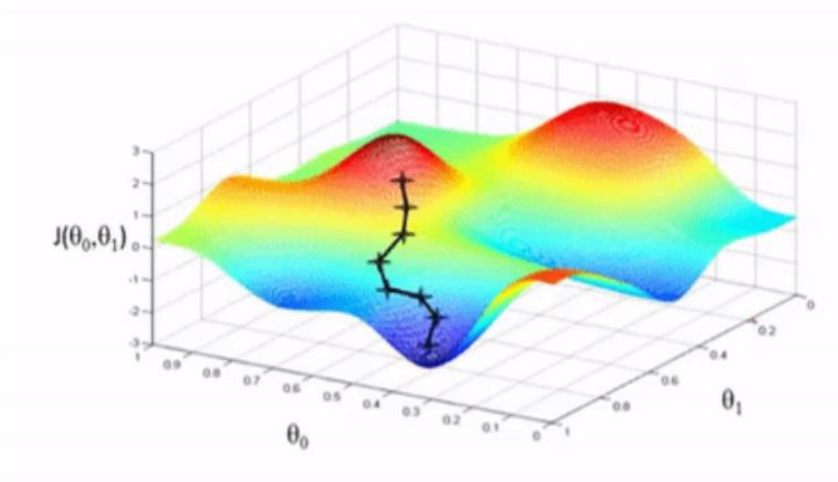
המטרה הראשית שלנו היא התכנסות כמה שיותר מהירה וכמה שיותר מדויקת למינימום הגלובלי.

כל האופטימיזרים משתמשים בגרדיאנט

(הנגזרת) ע"מ להגיע למטרה זו.

בין היתר גם קצב הלמידה של המודל הוא גורם

חשוב המשפיע לנו על ההתכנסות.



Learning rate

קיימות לנו כמה אופציות ע"מ לקבוע / לשפר קצב למידה למודל:

- לקבוע קצב למידה קבוע

- להשתמש באלגוריתמי תזמון לקצב הלמידה

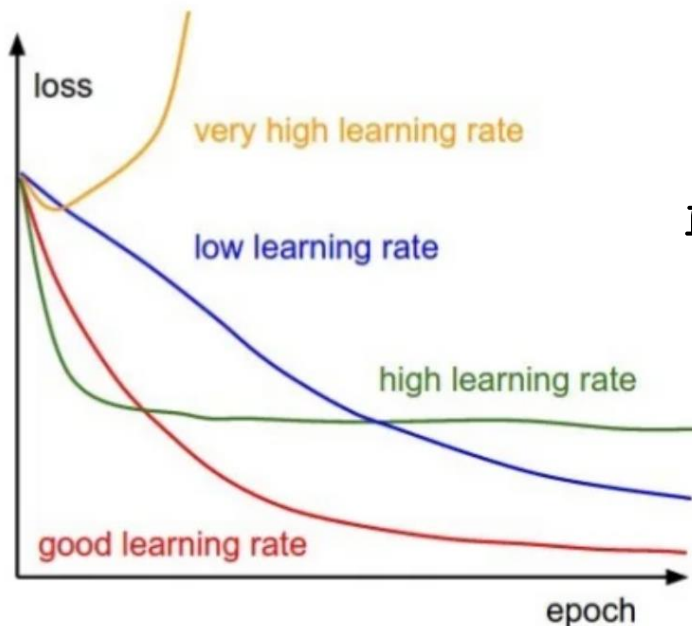
(Learning Rate Schedules) - שינוי לקצב הלמידה כפונקציה מסוימת

לפי מספר האיטרציה – קצב למידה אחיד לכל הפרמטרים.

להשתמש באופטימיזרים אדפטיבים – משנים את קצב הלמידה

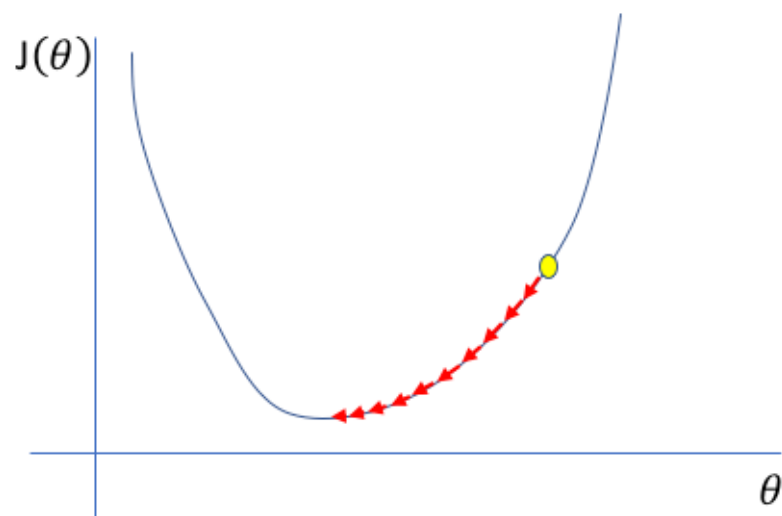
בהתאם לגרדיאנטים – קצב למידה שונה לכל פרמטר.

- Warm-up (נרחיב בהרצאה)



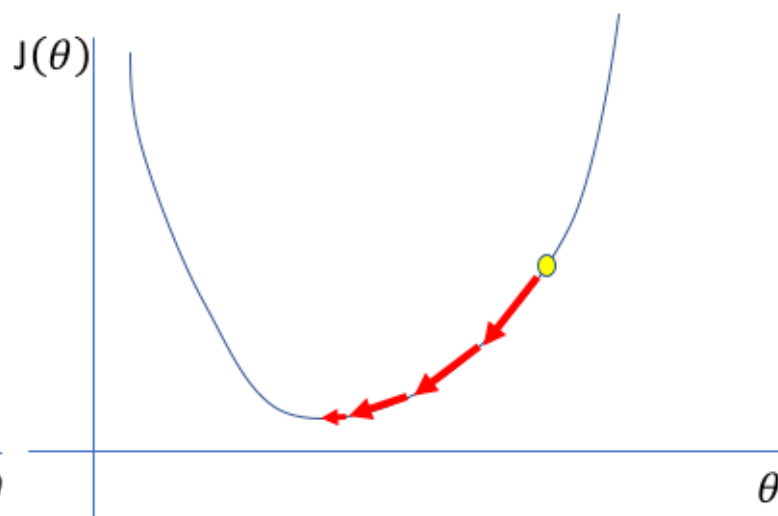
How the learning rate affects the process?

Too low



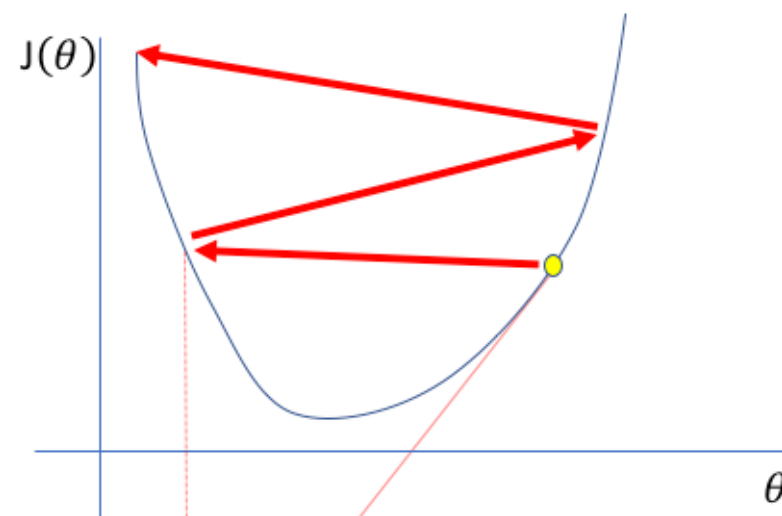
A small learning rate requires many updates before reaching the minimum point

Just right

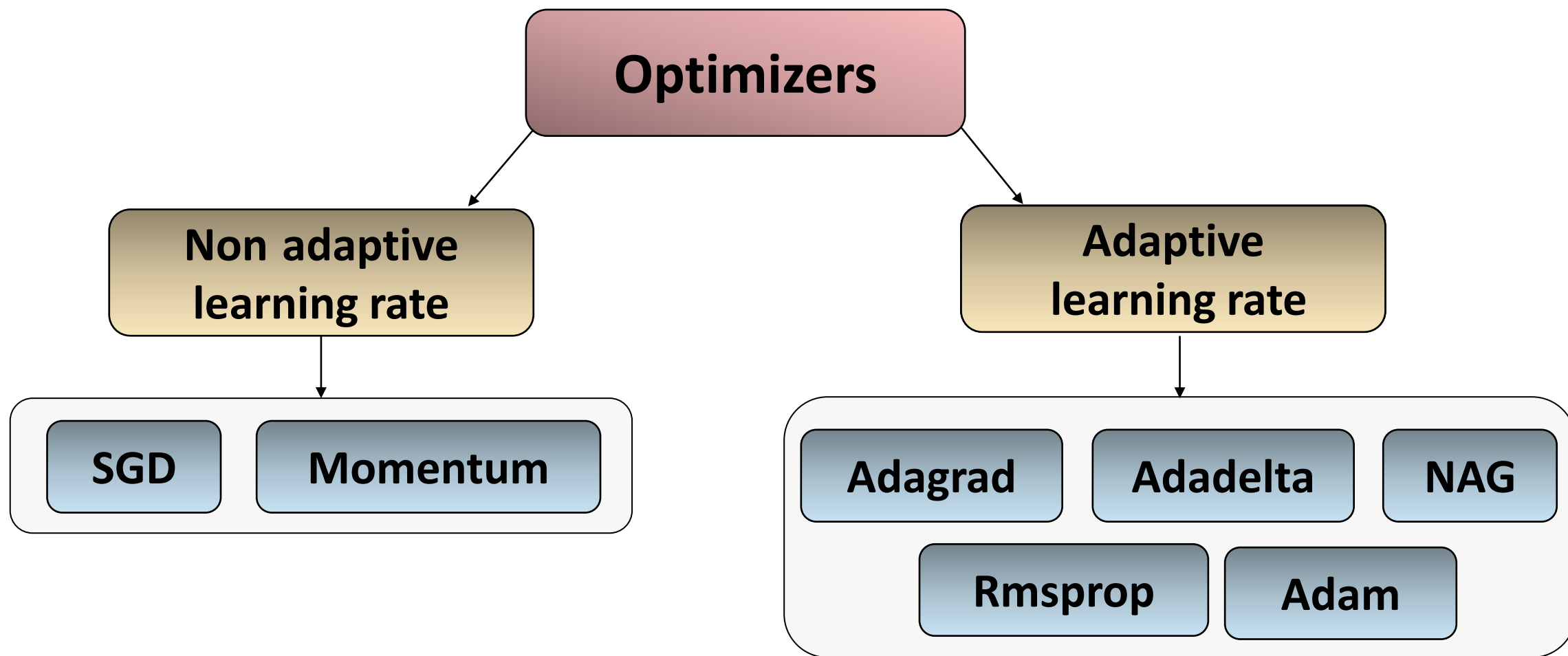
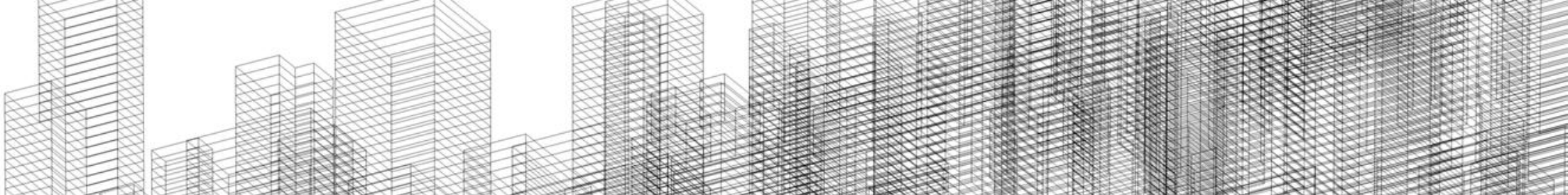


The optimal learning rate swiftly reaches the minimum point

Too high



Too large of a learning rate causes drastic updates which lead to divergent behaviors



Gradient descent

The target:

Reach the global
minimum

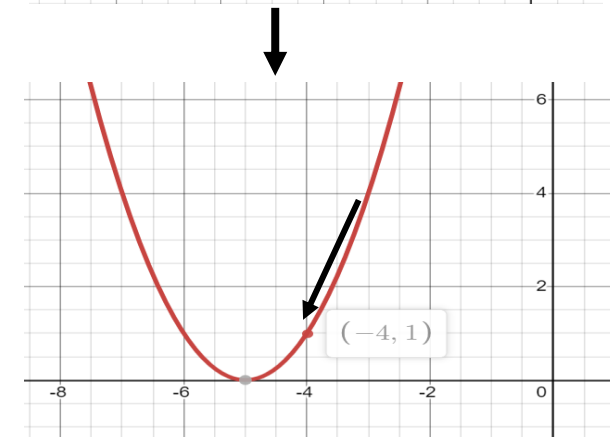
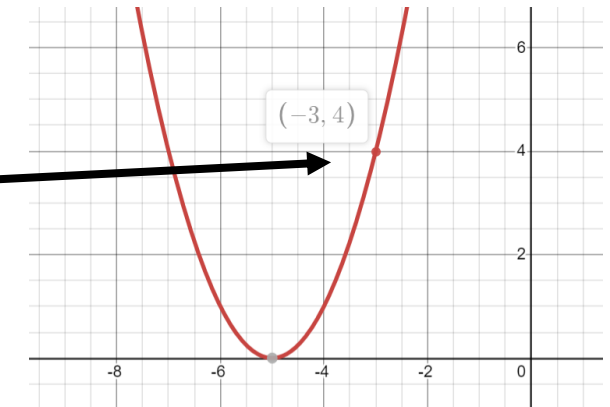
Random start



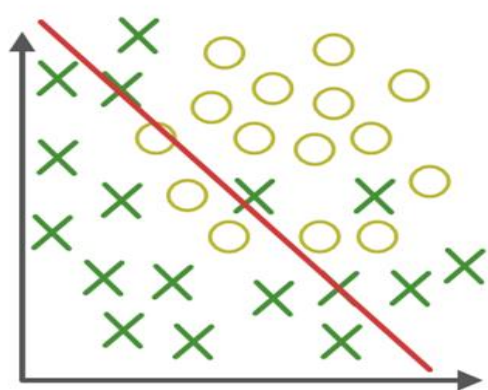
Computing the gradient



Move in the direction of
the **negative** of the **gradient**

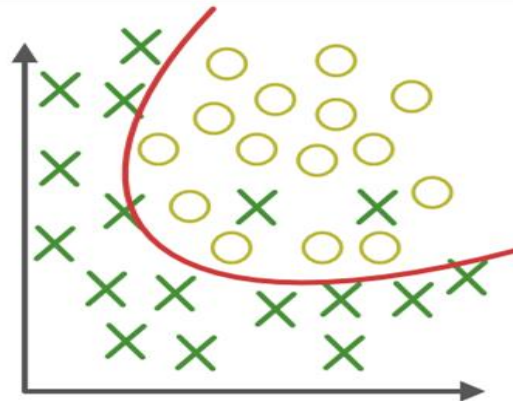


Fitting a model



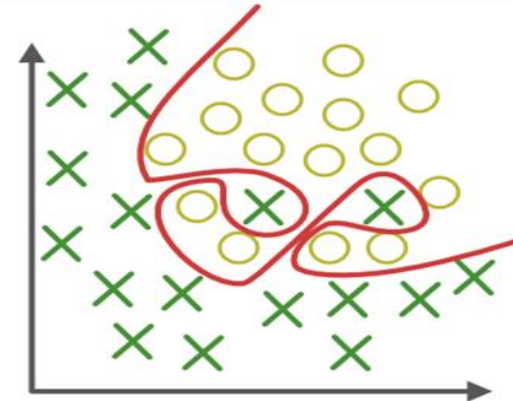
Under-fitting
(too simple to
explain the variance)

לרוב קורה כשהמודל שלנו
"פשוט מידי"



Appropriate-fitting

המצב האידיאלי
שהאופטימיזרים אמורים
לעזור לנו להגיע אליו.



Over-fitting
(forcefitting--too
good to be true)

כאשר יש לנו הצלחה גבוהה
בtrain והצלחה יותר נמוכה
בtest (בצורה משמעותית)