Gradient descent

גלעד מרקמן

Google Colab notebook

קריית החינוך
פארק המדע
בית לערכים
למצוינות ולחדשנות

חישוב מינימום בלמידת מכונה

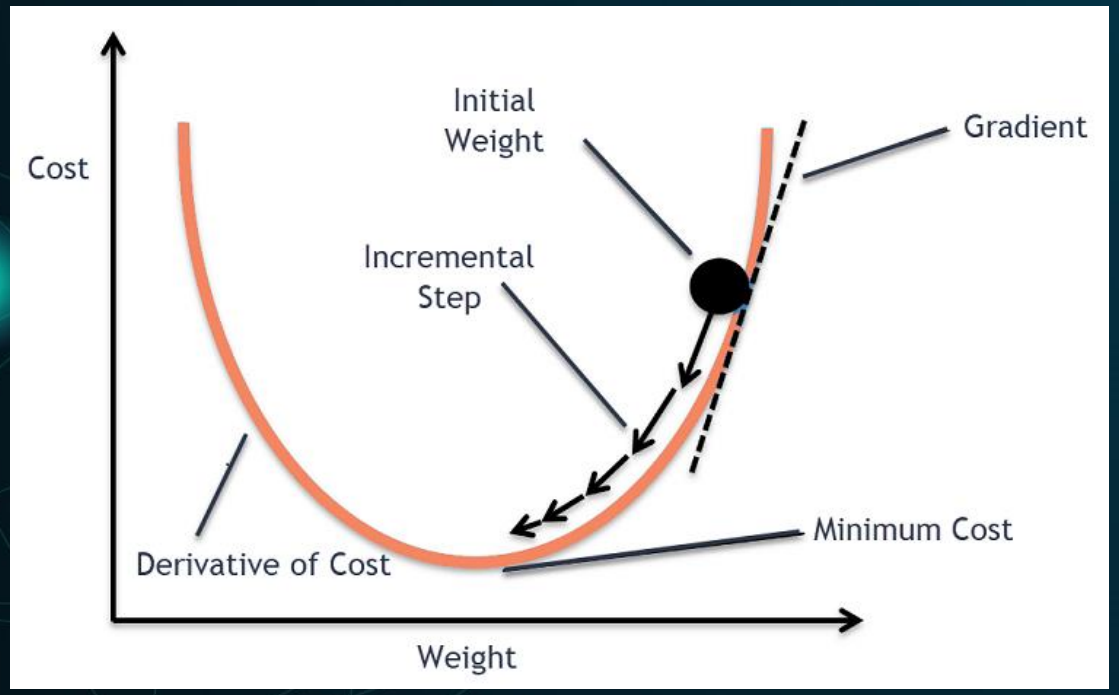
- באימון רשת נוירונים אנחנו מתחילים עם רשת נוירונים אקראית, ובונים פונקציה המחשבת את הטעות של הרשת שלנו.
- פונקציית הטעות, המכונה פונקציית המחיר lost() / cost(), מחשבת את הטעות של הרשת בהתבסס על הפרמטרים Weights של הרשת.
- באימון רשת הנוירונים אנחנו מחפשים את הפרמטרים שיקטינו את הטעות למינימום.
- למעשה אנחנו מחפשים את נקודת המינימום של פונקציית המחיר.

- לכן, יש חשיבות רבה לאלגוריתם שמוצא מינימום של פונקציה.

קריית החינוך
פארק המדע
בית לערכים
למצוינות ולחדשנות

קריית החינוך
פארק המדע
בית לערכים
למצוינות ולחדשנות

Gradient descent



- Gradient descent היא שיטה למציאת מינימום מקומי של פונקציה בעזרת חישוב הנגזרת.
- האלגוריתם מתחיל בנקודה שרירותית ומתקדם בצעדים קטנים בניגוד לכיוון הנגזרת. אם הנגזרת חיובית - מתקדמים לכיוון הירידה, ואם שלילית - מתקדמים לכיוון העליה.
- המטרה למצוא את המשקל weight המינימלי של פונקציה המחיר cost.
- בכל הדוגמאות שלנו:
  - ציר ה X יכונה משקל Weights או W.
  - ציר ה Y יכונה מחיר cost או lost.
- הערכים של ציר ה X יכונה גם פרמטרים.

מדוע להשתמש ב SGD בלמידת מכונה

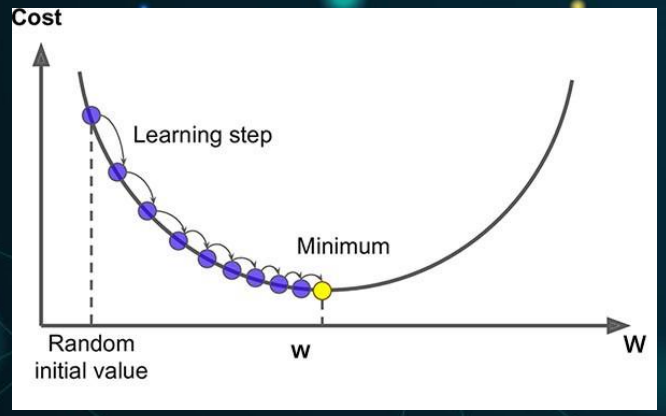
קריית החינוך
פארק המדע
בית לערכים
למצוינות ולחדשנות

- חישוב המינימום של פונקציות פשוטות יכול להיעשות בדרכים אריתמטיות. כגון: מציאת פונקציית הנגזרת, השוואתה ל- 0, ומציאת הפרמטר המאפס את פונקציית הנגזרת.
- בפונקציות מסובכות (כמו רשת נוירונים), עם פרמטרים רבים, אין דרך אלגברית למצוא את המינימום, ולכן ה Gradient Descent הינה השיטה המשמשת אותנו למציאת המינימום של פונקציית הטעות (המחיר) של רשת הנוירונים.
- בלמידת מכונה אנו נשתמש בסימונים הבאים (לפונקציה במימד אחד):
  - ציר ה X יכונה משקל Weight או W.
  - ציר ה Y יכונה מחיר cost או lost.
- כאשר יש פונקציה במספר מימדים / משתנים:
  - המשתנים/הפרמטרים יכונה משקלים w.
  - תוצאת החישוב תכונה cost או lost.

אלגורים SGD

קריית החינוך
פארק המדע
בית לערכים
למצוינות ולחדשנות

- האלגוריתם Stochastic Gradient Descent למציאת מינימום כולל את השלבים הבאים:
  - אתחול אקראי של המשקלים (בחירת נקודת ההתחלה, למשל w=0).
  - אתחול קצב הלמידה (learning_rate = 0.01)
  - לולאה:
    - חילחול קדימה – חישוב הפלט של הפונקציה בהתאם לפרמטרים.
    - חילחול לאחור - חישוב הנגזרת בהתאם לפרמטרים
    - עדכון הפרמטרים – על פי הנגזרת וקצב הלמידה.
    - איפוס הנגזרות.



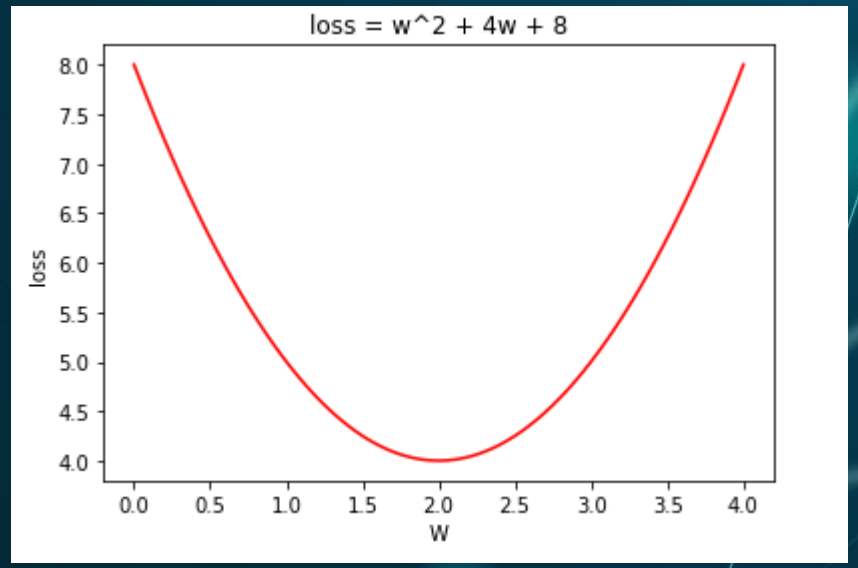
- עדכון הפרמטרים נעשה בניגוד לסימון הנגזרת לפי הנוסחה:
  - $w = w - grad * learning_rate$
- אם, לדוגמה, הנגזרת בנקודה w היא grad=2. הפונקציה עולה בנקודה זו ולכן אנחנו צריכים לנוע שמאלה (לרדת). נזיז את x שמאלה ב 2*0.01=0.02.

אלגוריתם למציאת מינימום

קריית החינוך
פארק המדע
בית לערכים
למצוינות ולחדשנות

- נדגים את האלגוריתם למציאת מינימום באמצעות פונקציה פשוטה:

- $loss = w^2 - 4w + 8$



- פונקציה זו הינה פונקציה פשוטה בה ניתן למצוא את המינימום גם בדרכים אלגבריות. אולם אנחנו נשתמש ב Gradient descent.

קריית החינוך
פארק המדע
בית לערכים
למצוינות ולחדשנות

הקוד לאלגוריתם SGD

- אתחול אקראי של הפרמטרים (נקודת ההתחלה)
- אתחול קצב הלמידה
- לולאה:
  - חילחול קדימה – חישוב הטעות
  - חילחול לאחור - חישוב הנגזרת
  - עדכון הפרמטרים בהתאם לנגזרת וקצב הלמידה.
  - איפוס הנגזרות.
- הדפסה של התוצאה.

```
# initialize random weight and params
W = torch.tensor(0.0, requires_grad=True)
learning_rate = 0.1

for epoch in range (100):
  # Forward
  loss = W**2-4*W+8

  # bacward - calculate grads
  loss.backward()

  # optimize wights
  with torch.no_grad():
    W -= learning_rate * W.grad

  if epoch % 10 == 0:
    print(f"epoch= {epoch} W= {W.item():.3f} model={loss:.3f}

  # zero wights
  W.grad.zero_()

print(f"End W= {W.item():.3f} model={loss:.3f} ")
```

```
epoch= 0 W= 0.400 model=8.000 grad= -4.000
epoch= 10 W= 1.828 model=4.046 grad= -0.429
epoch= 20 W= 1.982 model=4.001 grad= -0.046
epoch= 30 W= 1.998 model=4.000 grad= -0.005
epoch= 40 W= 2.000 model=4.000 grad= -0.001
epoch= 50 W= 2.000 model=4.000 grad= -0.000
epoch= 60 W= 2.000 model=4.000 grad= -0.000
epoch= 70 W= 2.000 model=4.000 grad= -0.000
epoch= 80 W= 2.000 model=4.000 grad= -0.000
epoch= 90 W= 2.000 model=4.000 grad= -0.000
End W= 2.000 model=4.000
```

קריית החינוך
פארק המדע
בית לערכים
למצוינות ולחדשנות

שימוש ב Optimizer

- לספריית autograd יש פונקציות מיוחדות שנועדו לעדכן את הפרמטרים בהתאם לתוצאות הנגזרת.
- השימוש ב Optimizer נעשה בשלבים הבאים:
  - איתחול ה optimizer. הפעולה מקבלת את הפרמטרים אותם יש לעדכן Weights ואת קצב הלמידה.

```
# init optimizer
optimizer = torch.optim.SGD([W], lr=learning_rate)
```

```
# optimize wights
optimizer.step()
```

- עדכון הפרמטרים נעשה באמצעות הפעולה ()step.

```
# zero wights
optimizer.zero_grad()
```

- איפוס הנגזרת נעשה באמצעות פעולה ()zero_grad

- בהמשך נכיר סוגים נוספים של אופטימזירים.

שימוש ב Optimizer

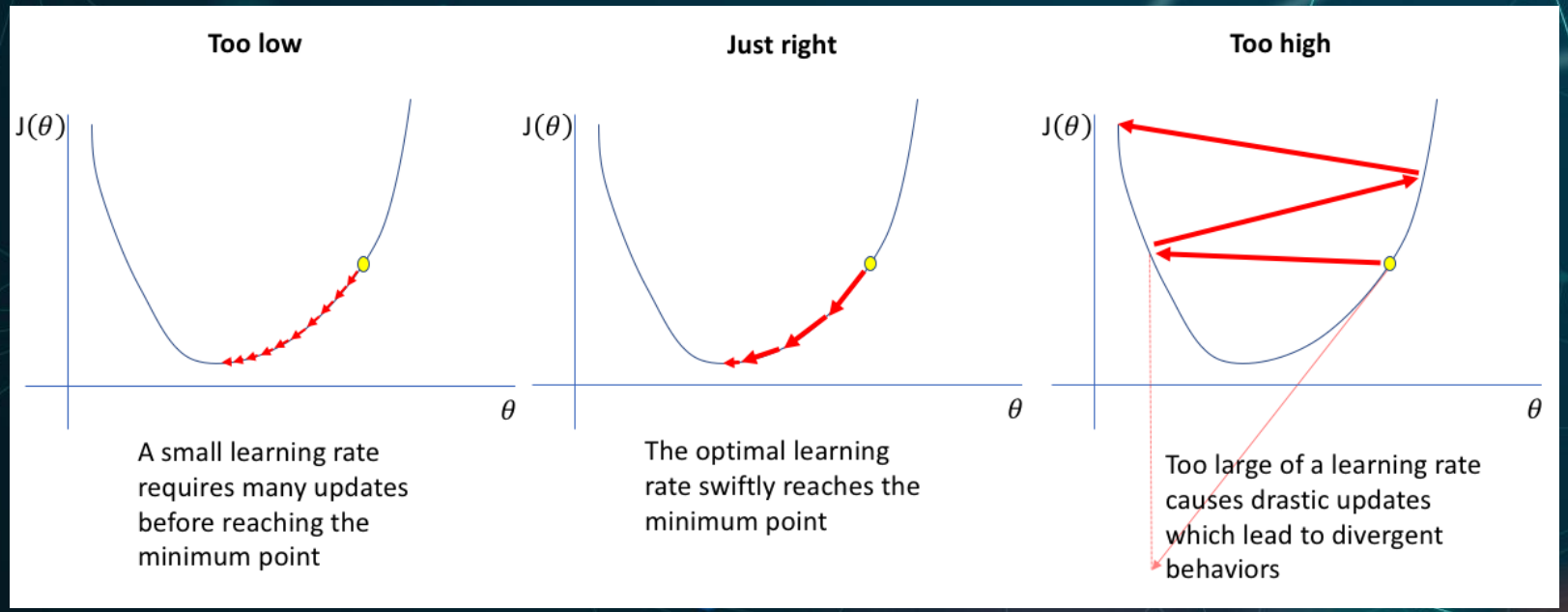
```python
# initialize random weight and params
W = torch.tensor(0.0, requires_grad=True)
learning_rate = 0.1
# init optimizer
optimizer = torch.optim.SGD([W], lr=learning_rate)

for epoch in range (100):
  # Forward
  loss = W**2-4*W+8

  # bacward - calculate grads
  loss.backward()

  # optimize wights
  optimizer.step()

  if epoch % 10 == 0:
    print(f"epoch= {epoch} W= {W.item():.3f} model={loss:.3f} grad= {W.grad:.3f}")

  # zero wights
  optimizer.zero_grad()

print(f"End W= {W.item():.3f} model={loss:.3f} ")
```

```
epoch= 0 W= 0.400 model=8.000 grad= -4.000
epoch= 10 W= 1.828 model=4.046 grad= -0.429
epoch= 20 W= 1.982 model=4.001 grad= -0.046
epoch= 30 W= 1.998 model=4.000 grad= -0.005
epoch= 40 W= 2.000 model=4.000 grad= -0.001
epoch= 50 W= 2.000 model=4.000 grad= -0.000
epoch= 60 W= 2.000 model=4.000 grad= -0.000
epoch= 70 W= 2.000 model=4.000 grad= -0.000
epoch= 80 W= 2.000 model=4.000 grad= -0.000
epoch= 90 W= 2.000 model=4.000 grad= -0.000
End W= 2.000 model=4.000
```

קריית החינוך פארק המדע
בית לערכים למצוינות ולחדשנות

Learning rate

- קצב הלמידה, יחד עם ערך הנגזרת, קובעים את הצעדים שנעשה בכל איטרציה לכיוון נקודת המינימום.
- קצב למידה נמוך מידי עלול לגרום לכך שנתקדם לאט מידי ולא נגיע למינימום. קצב למידה גדול מידי עלול לגרום לכך שנתבדר ולא נגיע כלל למינימום.
- אין כללים לקביעת קצב הלמידה. המדובר בניסוי וטעיה.



תרגיל

- נתונה הפונקציה $L = w^4 + 3w^3 - w^2 - 3w$
- כתוב תוכנית המחשבת את המינימום של הגרף באמצעות gradient descent. הרץ את התוכנית באמצעות שני הערכים התחלתיים הבאים:
  - w = 0
  - w = 2
  - w = -4
- הסבר מדוע קיבלת תוצאות שונות. מה משמעות התוצאות שקיבלת ?

קריית החינוך
פארק המדע
בית לערכים
למצוינות ולחדשנות