

ארכיטקטורת CNN (Convolutional Neural Network)

רקע: רשת ניורונית היא מבנה שמתיימר לדמות את המוח האנושי. לכל ניורון בשכבה יש קשר לניורונים בשכבות השכנות, ולכל קשר יש משקל. הניורונים "יורים פולסים", ועצמת הפולס שמגיעה לניורון היא קומבינציה לינארית בהתאם למשקלים של הפולסים מהניורונים המקושרים בשכבה הקודמת. את ערך הפלט מעבירים דרך פונקציה לא-לינארית (כמו סיגמואיד או טנגנס-היפרבולי או softmax) כדי לקבל תשובה דיסקרטית.

רשתות ניורוניות משמשות בדרך כלל לפתור בעיות של recognition classification - מצינים לרשת דוגמאות ולפי זה היא לומדת את המודל. אומרים לרשת "עבור קלט כזה הפלט המצופה הוא X " - למשל מראים לה תמונה של חתול ואומרים לה שעבור הקלט הזה (גריד של פיקסלים) הפלט המצופה הוא "חתול" (ניורון הפלט שמייצג חתול ידלק). התשובה מחושבת לפי המשקלים. נהוג לאתחל את הרשת עם משקלים אקראיים, ולכן בהתחלה היא תיתן תשובות שרירותיות ושגויות, אבל אפשר לחשב את פונקציית השגיאה ולעשות back propagation כדי להגיע לכלל לעדכון המשקלות השונים. מתבצע תהליך עדכון על הרבה דוגמאות ("epoch") והמשקלים מתכנסים ונותנים תשובות נכונות באחוזי דיוק יותר ויותר גבוהים.

במקרה של קונבולוציה, הקלט הוא תמונה בגודל למשל $3 \times 32 \times 32$ (גודל 32×32 ו-3 ערכי רמות אפור עבור כל צבע יסוד), ויש פילטר קונבולוציה בגודל למשל $3 \times 5 \times 5$. זה יוצר תמונה של $3 \times 28 \times 28$ (כי מאבדים את המסגרת בעובי 2 פיקסלים).

למסכה הזאת יש 75 מקדמים. לפני עידן למידת המכונה היינו צריכים לקבוע אותם ידנית - וזה יותר מדי בשביל בני אדם. עם רשתות ניורוניות, המקדמים של הפילטר הופכים למשקלי הקשרים ברשת. אם משתמשים בפילטרים אחרים אפשר לקבל שכבות ביניים אחרות שמוצאות דברים אחרים.

נשים ♡: לא צריך רשת עם קישוריות מלאה - מסתכלים לוקלית על אזורים קטנים בתמונה.

בסופו של דבר מגיעים לשכבות יותר ויותר fully-connected, עד שמגיעים לשכבת הפלט.

למה זה פתאום אפשרי?

- לימוד features אוטומטי
- כוח חישוב (GPU)
- STESATAD הרבה יותר גדולים (הרבה יותר דוגמאות)
- טכניקות וארכיטקטורות מתקדמות