CS231n: CNNs for Visual Recognition **Spring 2017**

Lecture 1 & 2

Course Coordinator: Prof. Fei-Fei Li *Scribes: Akash Gupta*

- **Image Classification:** A core problem in Computer Vision.
- **The Problem:** Semantic Gap
- **Challenges involved:**
  - Viewpoint variation
  - Illumination
  - Deformation
  - Occlusion
  - Background clutter
  - Intraclass variation

Data - Driven approach:

1. Collect a dataset of images and labels.
2. Use ML to train a classifier.
3. Evaluate the classifier on new/test images.

API has two components: train() and predict()

- train(): Memorize all data and labels.
- predict(): Predict the label of the most similar training image.

Image Classification example -

- Dataset: CIFAR10 (10 classes, 50k training images, 10k testing images)
- K- Nearest Neighbour Classifier: Fast at training time (O(1)) and Slow at testing time (O(N))
- Distance Metric: L1(Manhattan) distance - $d_1(I_1, I_2) = \sum_p |I_1^p - I_2^p|$
- Distance Metric: L2(Euclidean) distance - $d_1(I_1, I_2) = \sqrt{\sum_p (I_1^p - I_2^p)^2}$

  - L2 distance is invariant to rotation of the coordinate axes as contrary to L1.
- Hyperparameters: K, distance metric - Choices that we set rather than learn.

Q) *Where is L1 distance preferred over L2 distance?*
A) Best ans is it's mainly problem dependent so try both and see what works! L1 has coordinate dependency. Assuming a case that if individual elements in a vector have some meaning (like certain features) then L1 might be better.

- Setting hyperparameters -

- Best idea is to split the data in 3 sets: train, val and test. Choose hyperparameters on val and evaluate on test set.
- Cross - Validation: Split data into K folds and try each fold as validation set and average the results.

Q) *What if test data is not representative of the data out in the wild?*
A) Generally does not happen since dataset is curated from a same probability distribution. While splitting, always partition randomly to avoid these type of shifts.

- kNN never used for classifying images - very slow at test time, distance metrics on pixels are not informative, curse of dimensionality since kNN can only work if the training set is densely populated.

- Linear classification:

- Take an image(X) and set of parameters (W) and input it to a function $f(x, W) = Wx + b$ and output a vector of 10 scores with each score signifying the probability of the image to belong to the respective class.
- Similar to **template matching** where each row of the weight matrix is sorta learned template of the images in the training dataset.

Assignment-1

- to extract the maximum occurring element in an array -
  - counts = np.bincount(a)
    y_pred[i] = np.argmax(counts)
- Python broadcasting -
  - (4,) + (3,4) = (3,4)
  - (4,) + (4,3) - ERROR: Cannot broadcast together

Inline Questions:

Q) *Notice the structured patterns in the distance matrix, where some rows or columns are visible brighter. (Note that with the default color scheme black indicates low distances while white indicates high distances.)*

- What in the data is the cause behind the distinctly bright rows?
- What causes the columns?
- YourAnswer: The high value of the L2 distance calculated. Outliers in the dataset.

Q) *Which of the following preprocessing steps will not change the performance of a Nearest Neighbor classifier that uses L1 distance? Select all that apply.*

- Subtracting the mean $\mu(\tilde{p}_{ij}^{(k)} = p_{ij}^{(k)} - \mu)$
- Subtracting the per pixel mean $\mu_{ij}(\tilde{p}_{ij}^{(k)} = p_{ij}^{(k)} - \mu_{ij})$
- Subtracting the mean μ and dividing by the standard deviation σ.
- Subtracting the pixel-wise mean μ_{ij} and dividing by the pixel-wise standard deviation σ_{ij}.
- Rotating the coordinate axes of the data.
- YourAnswer: 1,2,3,4
- YourExplanation: 1)No change on subtracting the mean as all pixels are shifted by the same amount in data space.

  2) Same with standard as all shifted points are scaled by same amount so distance also scaled by same amount. So biggest still remains biggest.

  3) Rotating does change performance as relative distance does not remains the same.