

1. Machine Learning adalah suatu computer atau program computer yang mengalami peningkatan seiring pembelajaran pada contoh ataupun data set sehingga developer tidak perlu susah dalam coding pemrograman.
2. SAVMAP <http://lasig.epfl.ch/savmap> . Salah satu objective pada SAVMAP adalah classification hewan semi-arid savanna di Namibia dengan image processing yang diambil dengan Unmanned Aircraft Vehicle (UAV). Mereka mencoba beberapa metode machine learning classifier diantaranya LOG , SVM with an RBF Kernel, NB, dan RF dan mencoba mengcomparasinya dengan hasil performasi 80% presisi dan recall menggunakan HOG features untuk feature set.
3. ~
 - A. 0 – 1 jika input bukan non negative dan (-1 – 1) jika input negative.
 - B. Ya sama, karena range nilai yang terdapat pada similarity adalah dari 0 – 1 dimana 0 menyatakan kedua data tidak mirip dan semakin besar tingkat kemiripan akan menyatakan data tersebut sama.
 - C. Jika kita kurangkan rata-rata dari x dan y maka pengukuran menjadi invariant dan hasilnya adalah titik tengah (Pearson Product Moment Correlation Coefficient) PPMC dimana sudutnya adalah pusat dari vector.
 - D. Euclidean :

$$d_{euclid}(p, q) = |p - q| = \sqrt{\sum_i (p_i - q_i)^2}$$
$$d_{euclid}(p, q) = \sqrt{\sum_i p_i^2 - 2 \sum_i p_i q_i + \sum_i q_i^2}$$

Jika panjang = 1, maka jaraknya ke 0 adalah

$$|p| = d_{euclid}(p, 0) = \sqrt{\sum_i (p_i - 0)^2} = \sqrt{\sum_i p_i^2}$$

Jika hasilnya 1 maka kuadratnya juga 1: $\sqrt{\sum_i p_i^2} = 1$

Kemudian pada rumus Euclid yang diekspansi, selanjutnya substitusi

$$d_{euclid}(p, q) = \sqrt{1 - 2 \sum_i p_i q_i + 1} = \sqrt{2 - 2 \sum_i p_i q_i}$$
$$= \sqrt{2(1 - \sum_i p_i q_i)}$$

Cosine Similarity :

$$s_{\cosine}(p, q) = \frac{p \cdot q}{|p||q|} = \frac{\sum_i p_i q_i}{|p||q|}$$

Jika panjang norm = 1 maka

$$s_{\cosine}(p, q) = \sum_i p_i q_i$$

Karena nilai cosine similarity selalu diantara 0 dan 1, maka kita bisa defenisikan jarak cosine dengan :

$$d_{\cosine}(p, q) = 1 - \sum_i p_i q_i$$

Misalnya kita punya vector v dan w dengan menggunakan euclid, dan v lebih dekat ke suatu point p daripada w.

$$\begin{aligned} d_{euclid}(p, v) &\leq d_{euclid}(p, w) \\ \sqrt{2(1 - \sum_i p_i v_i)} &\leq \sqrt{2(1 - \sum_i p_i w_i)} \\ 2(1 - \sum_i p_i v_i) &\leq 2(1 - \sum_i p_i w_i) \\ (1 - \sum_i p_i v_i) &\leq (1 - \sum_i p_i w_i) \end{aligned}$$

Sehingga, kita dapat hubungan antara cosine similarity dan Euclidean

$$d_{euclid}(p, v) \leq d_{euclid}(p, w) \approx d_{\cosine}(p, v) \leq d_{\cosine}(p, w)$$

E. Euclidean

$$d_{euclid}(x, y) = \sqrt{\sum_i x_i^2 - 2 \sum_i x_i y_i + \sum_i y_i^2}$$

Correlation

$$r(x, y) = \frac{\frac{1}{n} \sum_i x_i y_i - \mu_x \mu_y}{\sigma_x \sigma_y}$$

jika X dan Y distandarkan, mereka masing-masing akan memiliki rata-rata 0 dan standar deviasi 1, sehingga rumus menjadi

$$r(x, y) = \frac{1}{n} \sum_i x_i y_i$$

Kembali ke rumus Euclidean jika x dan y distandarkan maka keduanya = n. Sisa $\sum xy$ sebagai non-konstan, maka formula bisa disederhanakan menjadi :

$$r(x, y) = 1 - \frac{d_{euclid}^2(x, y)}{2n}$$

4. ~

- A. Pertama berdasarkan pairwise proximity bisa menggunakan minimum pairwise similiarity atau maximum pairwise dissimilarity. Kedua berdasarkan data di Euclidean kemudian hitung total atau rata-rata jarak data ke centroid.
- B. Dengan menghitung jarak antara centroids dari kedua kumpulan data.
- C. Dengan menghitung rata-rata kedekatan pasangan dari kumpulan data di satu kelompok group dengan kumpulan data lain. Atau dengan mengambil kedekatan yang maksimal atau minimum.

5. ~

- A. Untuk satu document mempunyai bobot maximum yaitu log m.
- B. Kata yang muncul pada setiap document memiliki bobot 0
- C. Pengamatan dari kata yang ada di setiap document tidak mempengaruhi document lain.
- D. Near Duplicate Image Detection: min-Hash and tf-idf Weighting
<https://www.robots.ox.ac.uk/~vgg/publications/papers/chum08a.pdf>