7- PCA and SVMs

measure: distance is called measure when:

① symmetry = $d(\theta_1, \theta_2) = d(\theta_2, \theta_1)$

② self similarity = $d(\theta_1, \theta_1) = 0$

③ positivity = $d(\theta_1, \theta_2) = 0 \iff \theta_1 = \theta_2$

☆ a measure is called metric, if and only if it holds the triangular inequality

↳ $d(\theta_1, \theta_3) \leq d(\theta_1, \theta_2) + d(\theta_2, \theta_3)$

minkowski metric = $d(x, y) = \left(\sum_{i=1}^{\ell} |x_i - y_i|^q \right)^{\frac{1}{q}}$ (ℓ_1 = manhattan, ℓ_2 = euclidean, ℓ_∞ = tchebychev $= \max(|x_i - y_i|)$)

(L_q norm)

↳ as the number of dimension increases, minkowski distance becomes meaningless

reducing dimentionality =

↳ feature selection based on domain knowledge (eliminatin redundant and irrelevant features)

↳ gain-based feature selection (trying different subsets and taking the best one)

↳ distribution based feature selection (principle component analysis)

$cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - E(X).E(Y)$

covariance matrix = $\Sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)]$

mahalanobis distance = $\sqrt{(\vec{x}-\vec{y})^T \Sigma^{-1} (\vec{x}-\vec{y})}$ (distance between a point and a distribution)

↳ distance of the test point from the center of mass divided by the width of the ellipsoid in the direction of the test point

↳ unlike euclidean, it also seeks to measure the correlation between variables

☆ if all variables are independent, mahalanobis dist degrades into normalized euclidean dist.

principle component analysis (PCA) = technique that transforms high-dimensional data into lower-dim while preserving as much info using dependencies between variables.

w = direction of first principle component (unit vector)

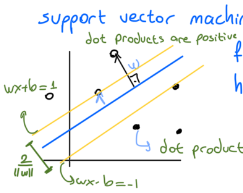
variance = of the data along the direction = $w^T \underbrace{\Sigma}_{\text{covariance matrix}} w$

☆ independent dimensions have largest variance

↳ $\Sigma w = \lambda w \longrightarrow \lambda = w^{-1} \Sigma w$ ☆ highest eigenvalue is the first principle component

- each succeding component is orthogonal to the previous ones

support vector machines (SVMs) = the boundary hyperplane can have infinite dimensions



dot products with w will be negative

$f(x) = w^T x + b$

$\underbrace{h(x)}_{\text{model}} = \begin{cases} -1, & \text{if } w^T x + b < 0 \quad \leq -1 \\ 1, & \text{if } w^T x + b \geq 0 \quad \geq 1 \end{cases}$

boundary = $w^T x + b = 0$

margins = $w^T x + b = \pm 1$

↳ $y^{(i)} . (w^T x^{(i)} + b) \geq 1$ (short form)

maximize $\left(\frac{2}{||w||} \right) \longrightarrow$ minimize $(||w||) \approx$ minimize $\left(\frac{1}{2} ||w||^2 \right)$

☆ $\min_{w,b} \max_{\alpha \geq 0} \left\{ \frac{1}{2} ||w^2|| - \sum_{i=1}^{n} \alpha_i [y^{(i)} (w.x^{(i)} + b) - 1] \right\}$

α_i = lagrange multipliers (nonzero only for support vectors)

↳ $w = \sum_{i=1}^{n} \alpha_i \, y^{(i)} . x^{(i)}$

↳ $b = \frac{1}{|SV|} \sum_{i \in SV} w.x^{(i)} - y^{(i)}$

SV = support vectors' set ($\alpha_i > 0$)

soft margins = allows misclassification