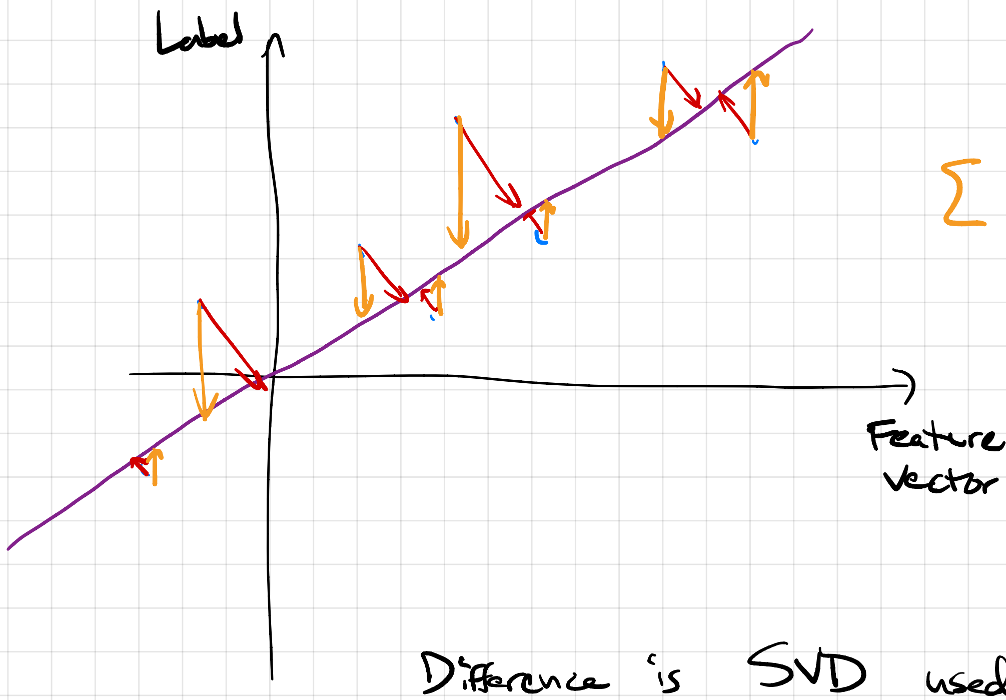
18 Feb 2022

SVD

Comparing SVD with Least Squares Regression.



$\sum \text{red}^2$ is SVD min. obj.

$\sum \text{orange}^2$ is LSQ min obj.

Difference is SVD used for unsupervised learning (identifying structure in data)

LSQ used for regression (approximating data labels using a learned function)

CLAIM: the optimal $v_1, \ldots, v_k$ can be found by solving

1. $v_1 \in \arg\max \{ \|Av\|_2^2 \mid \|v\|_2 = 1 \}$
2. $v_2 \in \arg\max \{ \|Av\|_2^2 \mid \|v\|_2 = 1, \langle v_1, v \rangle = 0 \}$

$\vdots$

k. $v_k \in \arg\max \{ \|Av\|_2^2 \mid \|v\|_2 = 1, \langle v_i, v \rangle = 0 \ \forall i < k \}$

Lemma. If X is an inner product space, W_1, W_2 are linear subspaces of X
(nonempty subset closed under $+, \cdot$)
and $\dim W_1 > \dim W_2$ then
$\exists\, w_1 \in W_1$ s.t. $\langle w_1, w_2 \rangle = 0 \quad \forall w_2 \in W_2$
and $w_1 \neq 0$.

Proof. Look at $T: W_1 \longrightarrow W_2^*$ defined by
$T(w_1) =$ the $f \in W_2^*$ such that $f(w_2) = \langle w_1, w_2 \rangle$.
If $w_1, \ldots, w_k$ are a basis of W_1
then $T(w_1), \ldots, T(w_k)$ are vectors in W_2^*
whose dimension satisfies $\dim(W_2^*) = \dim(W_2) < k$.
k vectors in a v.s. of dimension $< k$ are linearly dependent: $\exists\, a_1, \ldots, a_k$ not all zero
s.t. $a_1 T(w_1) + \ldots + a_k T(w_k) = 0 \in W_2^*$.

Let $w = a_1 w_1 + \cdots + a_k w_k$. $T(w) = 0 \in W_2^*$.
$\Downarrow$
$w \neq 0$ because
$w_1, \ldots, w_k$ are a basis
$a_1, \ldots, a_k$ not all zero.

$\exists\, w' \in W_2$, $\langle w, w' \rangle = \langle 0, w' \rangle = 0$.

Proving greedy algo works. Inductive hypothesis is that
$V' = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_{k-1} \\ | & & | \end{bmatrix}$ maximizes $\|AV'\|_F^2$ subj. to
orthonormality constraints.

Base case $k=1$: proven Wednesday.

For $k>1$: suppose $W = \begin{bmatrix} | & | & & | \\ w_1 & w_2 & \cdots & w_k \\ | & | & & | \end{bmatrix}$

maximizes $\|AW\|_F^2$ subj. to orthonormality.

We seek to compare $\|AW\|_F^2$ with $\|AV\|_F^2$ where V is the k-column matrix found by greedy alg.

Let W_1 = col space of W

W_2 = col space of V' (first $k-1$ columns of V)

Apply Lemma above: $\exists\, w \in W_1$, $w \neq 0$, s.t. $\langle w, v_i \rangle = 0$ for $i = 1, \ldots, k-1$.

WLOG $\|w\|_2 = 1$.

$\boxed{\text{WLOG} \quad w = w_k.}$

$\exists$ an orthormal basis of W_1, $\begin{bmatrix} | & & | \\ w_1' & \cdots & w_k' \\ | & & | \end{bmatrix} = W'$

such that $w_k' = w$.

$W' = WQ$ for some $k \times k$ orthogonal matrix Q.

Recall $\|M\|_F^2 = \sum_{i,j} M_{ij}^2 = \sum_{i,j} M_{ij} (M^T)_{ji}$

$= \sum_i (MM^T)_{ii} = Tr(MM^T)$.

$$\|AW'\|_F^2 = \|AWQ\|_F^2 = Tr(AW\underbrace{QQ^T}_{=I}W^TA^T)$$

$$= Tr(AWW^TA^T) = \|AW\|_F^2.$$

Since $\|w\|_2 = 1$ and $\langle w, v_i \rangle = 0 \quad \forall i < k$,

greedy chose v_k with $\|Av_k\|_2^2 \geq \|Aw\|_2^2$.

$$\|Aw\|_2^2 = \|Aw_k\|_2^2.$$

Induct hypothesis:

$$\sum_{i<k} \|Av_i\|_2^2 \geq \sum_{i<k} \|Aw_i\|_2^2.$$

Combine 2 inequalities: $\sum_{i=1}^{k} \|Av_i\|_2^2 \geq \sum_{i=1}^{k} \|Aw_i\|_2^2$

i.e. $\|AV\|_F^2 \geq \|AW\|_F^2$.

How to solve $\max \left\{ \|Av\|_2^2 \;\middle|\; \|v\|_2^2 = 1, \langle v, v_i \rangle = 0 \; \forall i < k \right\}$?

<u>Fact.</u> Any symmetric matrix B can be written as $B = QDQ^T$ such that

- Q is orthogonal
- D is diagonal
- diagonal entries $d_{ii} = \lambda_i$ are the eigenvalues of B
- columns of Q, Qe_i are the corresponding eigenvectors.

$$\max \ \|Av\|_2^2 \quad \text{s.t.} \ \|v\|_2 = 1 \qquad \equiv \qquad \max \ (Av)^T Av = v^T \underbrace{A^T A}_{B} v \quad \text{s.t.} \ \|v\|_2 = 1.$$

B is square & symmetric $\Rightarrow$ $B = QDQ^T$

$$\max \ v^T Q D Q^T v \quad \text{s.t.} \ \|v\|_2 = 1 \qquad w = Q^T v$$

$$\Downarrow$$

$$\max \ w^T D w = \sum \lambda_i w_i^2$$

$$\text{s.t.} \ \|w\|_2 = 1. \qquad \text{s.t.} \ \sum w_i^2 = 1.$$

Opt. is $\max\{\lambda_i\}$ attained when $w = e_i$

$v = Qw = Qe_i$.

Right singular vectors of A.

Punchline: $v_1, \dots, v_k$ are the eigenvectors of $B = A^T A$ corresponding to its k largest eigenvalues.