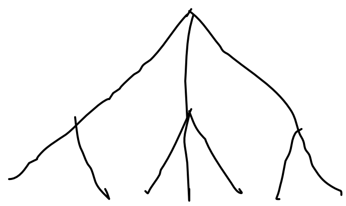
## Generating functions (continuing from last class)

Sequence of probabilities: $P_0$, $P_1$, $P_2$, …

let $f(x) = \sum_{i=0}^{\infty} P_i x^i$, so $f(1) = 1$, $f'(x) = \sum_{i=0}^{\infty} iP_i x^i$, $E[x] = \sum_{i=0}^{\infty} iP_i = f'(1)$

$$f^2(x) = P_0P_0x^0 + (P_0P_1 + P_1P_0)x^1 + \cdots$$

$$f^n(x) = \cdots$$

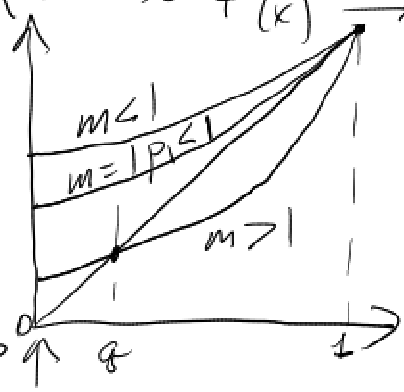


Let $f(x)$ be the function for the number of children that a node has.

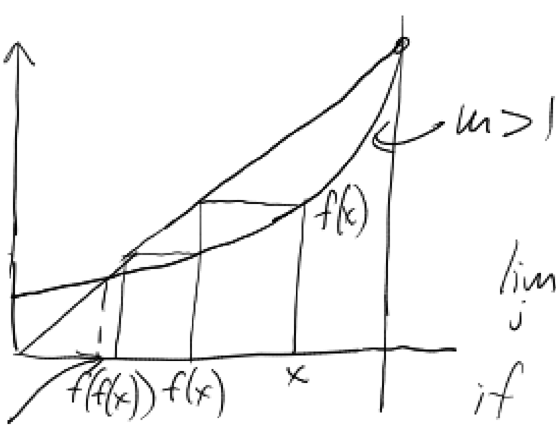
Here, $f_1(x) = f(x)$, and for each node in each row j, $f_{j+1}(x) = f_j(f(x))$.

$\Lambda + \Lambda + \Lambda + \ldots \sim f^i(x) \rightarrow$ replace x by f(x)

$f(x)=x$




intercept is $p_0$




$\lim_j f_j(x) = q$

if $f_j(x) = q$, $f_j(x) = q + 0x + 0x^2 + \ldots$

q Extinction probability

$$\lim_j f_j(x) \rightarrow q$$

Extinction probability q, immortality probability 1-q

## Singular Value Decomposition (SVD)

diagonal matrix – positive reals

$$\underset{\sim}{A} = \underset{\sim}{U}\,\underset{\sim}{D}\,\underset{\sim}{V}^T$$

columns are orthonormal

$$\begin{pmatrix} - & & & & & & \\ & - & & & & & \\ & & - & & & & \\ & & & - & & & \\ & & & & 0 & & \\ & & & & & 0 & \\ & & & & & & 0 \end{pmatrix}$$

rank

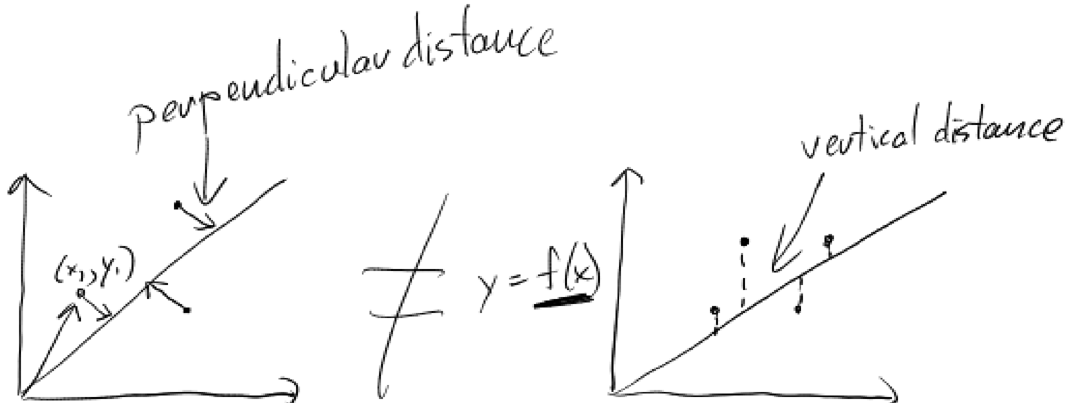
Because of roundoff, usually rank = size, but consider "small" values = 0

Where U and V have unit orthonormal columns (dot product of zero, magnitude of 1), and D has positive real values along the diagonal, and zeroes elsewhere.

Given matrix A is [n x d] dimensions. A can be thought of as n concatenated d-dimensional vectors.

The Frobenius norm of A is $\sum_{i,j} a_{ij}^2$. (sum of the squares of each element).

This can be used in a different type of linear fitting: normally, for least squares fitting we minimize the squares of the vertical distance to a line. Frobenius least squares fitting minimizes the perpendicular distance to the line instead.




i.e. Finding best 1D subspace to represent the data

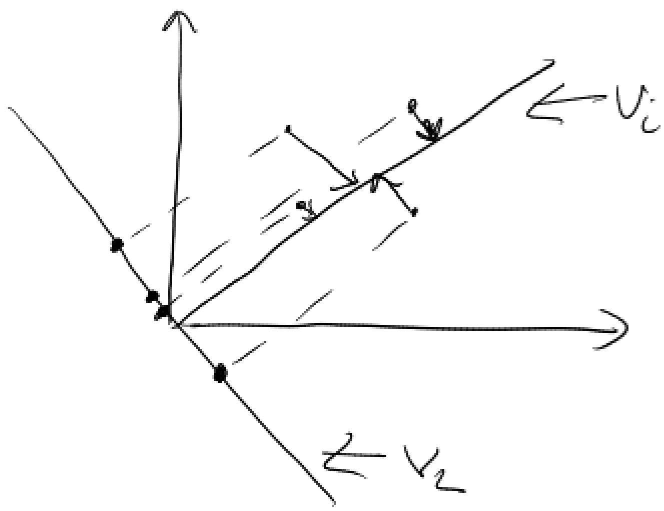
For a given point $(x_i, y_i)$, ${x_i}^2 + y_i^2 = (length\ of\ projection)^2 + (distance)^2$

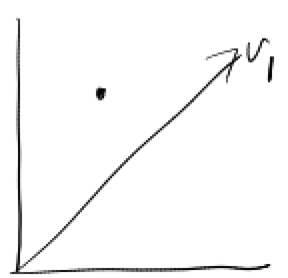
We want to use: $(distance)^2 = x_i^2 + y_i^2 - (length\ of\ projection)^2$.
Now, we can see that minimizing the distance is the same as maximizing the length of the projection!

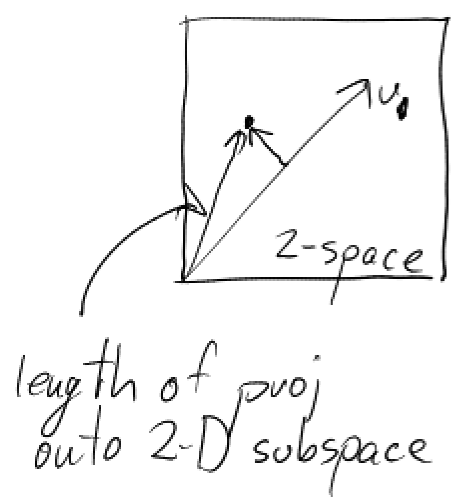
For a singular vector:
Let *v* bee the unit vector along the best fit line. The length of the projection of the $i^{th}$row of $A$, $a_i$, onto *v* is $|a_i\ .\ v|$. The sum of the length of squares of projections is: $\sum_i |a_i . v|^2 = |Av|^2$.

Define first singular vector $v_1$ of $A$ to be unit vector along best fit line through the origin for n points which are the rows of $A$. This $v_i = avg\ \max|Av|^2$, $|v| = 1$, value $\sigma_1 = |Av_1|$ is called the first singular value.

Given that matrix A is [n x d], we can look at it as a dimensionality-reduction problem: we are now looking at it like:




Greedy approach: To find the best 2-dimensional subspace of A, we find $v_1$, and then find the best 2-d subspace of $v_1$.