<u>Recall</u> n measurements of m variables recorded as vectors $x_1, \ldots, x_n \in \mathbb{R}^m$ have mean

$$\mu = \frac{1}{n}(x_1 + \cdots + x_n),$$

mean-centered data matrix $B \in M_{m \times n}(\mathbb{R})$ with i-th column $x_i - \mu$ and covariance matrix

$$S = \frac{1}{n-1} BB^T \in M_{m \times m}(\mathbb{R}).$$

S is symmetric, so the spectral theorem (and corollary on matrices BB^T) imply that S has nonnegative real eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0.$$

Let $u_1, \ldots, u_m$ be the corresponding orthogonal eigen~~values~~vectors.

The vectors $u_1, \ldots, u_m$ are the <u>principal components</u> of the data.

<u>Note</u> Total variance $T = Tr(S) = \lambda_1 + \cdots + \lambda_m$.

- The direction (unit vector) u_1 (the first principal direction) accounts for $\frac{\lambda_1}{T}$ of the total variance. The second principal direction u_2 accounts for $\frac{\lambda_2}{T}$ of the total variance. Etc.
- Thus u_1 points in the "most significant" direction of the data set.
- Amongst $u_1^\perp$, u_2 points in the most significant direction. Etc.

<u>Fact</u> The line spanned by u_1 minimizes <u>orthogonal</u> distance from line to point cloud (compare to least squares).

Suppose we are measuring 10 variables ~~and the~~ $T = 100$, $\lambda_1 = 90.5$, $\lambda_2 = 8.9$. Then $\lambda_3, \ldots, \lambda_{10} \leq 0.1$ and the data set in $\mathbb{R}^{10}$ has 99.4% of its total variance explained by $\mathrm{span}\{u_1, u_2\}$.