lets assume we have m data points $x^1, \dots, x^m \in R^d$
and the mean is $\mu = \frac{1}{m} \sum_{i=1}^{m} x^i$.

We will find direction w in the space R^d in which
$\|w\| \leq 1$ and it captures the maximum
variance. Therefore:

$$\max_{w: \|w\| \leq 1} \underbrace{\frac{1}{m} \sum_{i=1}^{m} (w^T x^i - w^T \mu)^2}_{\text{maximize it.}}$$

$$\frac{1}{m} \sum_{i=1}^{m} (w^T x^i - w^T \mu)^2 = \frac{1}{m} \sum_{i=1}^{m} (w^T (x^i - \mu))^2$$

$$= \frac{1}{m} \sum_{i=1}^{m} w^T (x^i - \mu)(x^i - \mu)^T w$$

$$= w^T \left(\frac{1}{m} \sum_{i=1}^{m} (x^i - \mu)(x^i - \mu)^T \right) w = w^T \underset{\text{covariance}}{C} w$$

the optimization $\Longrightarrow$ $\boxed{\max_{w: \|w\| \leq 1} w^T C w}$