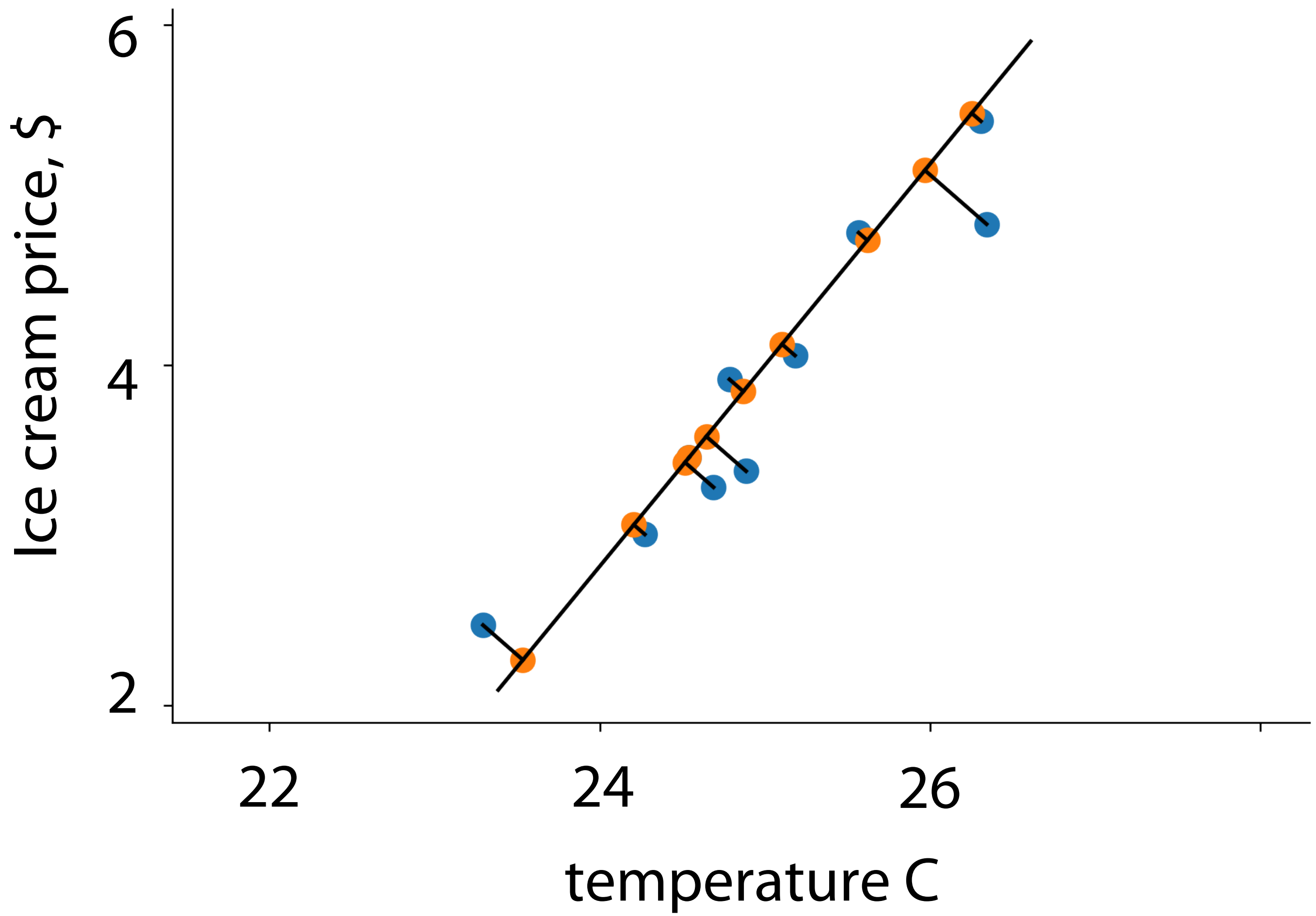
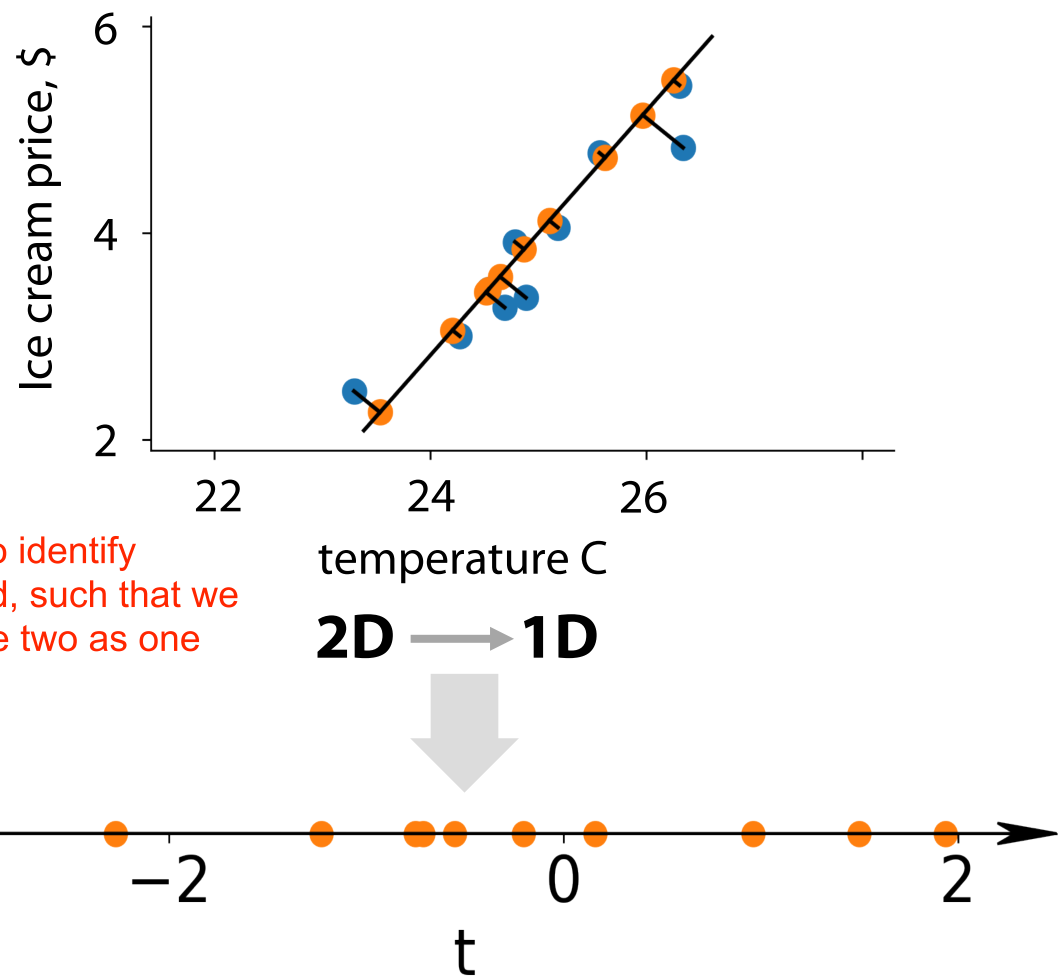
Ice Cream conspiracy

Principal Component Analysis

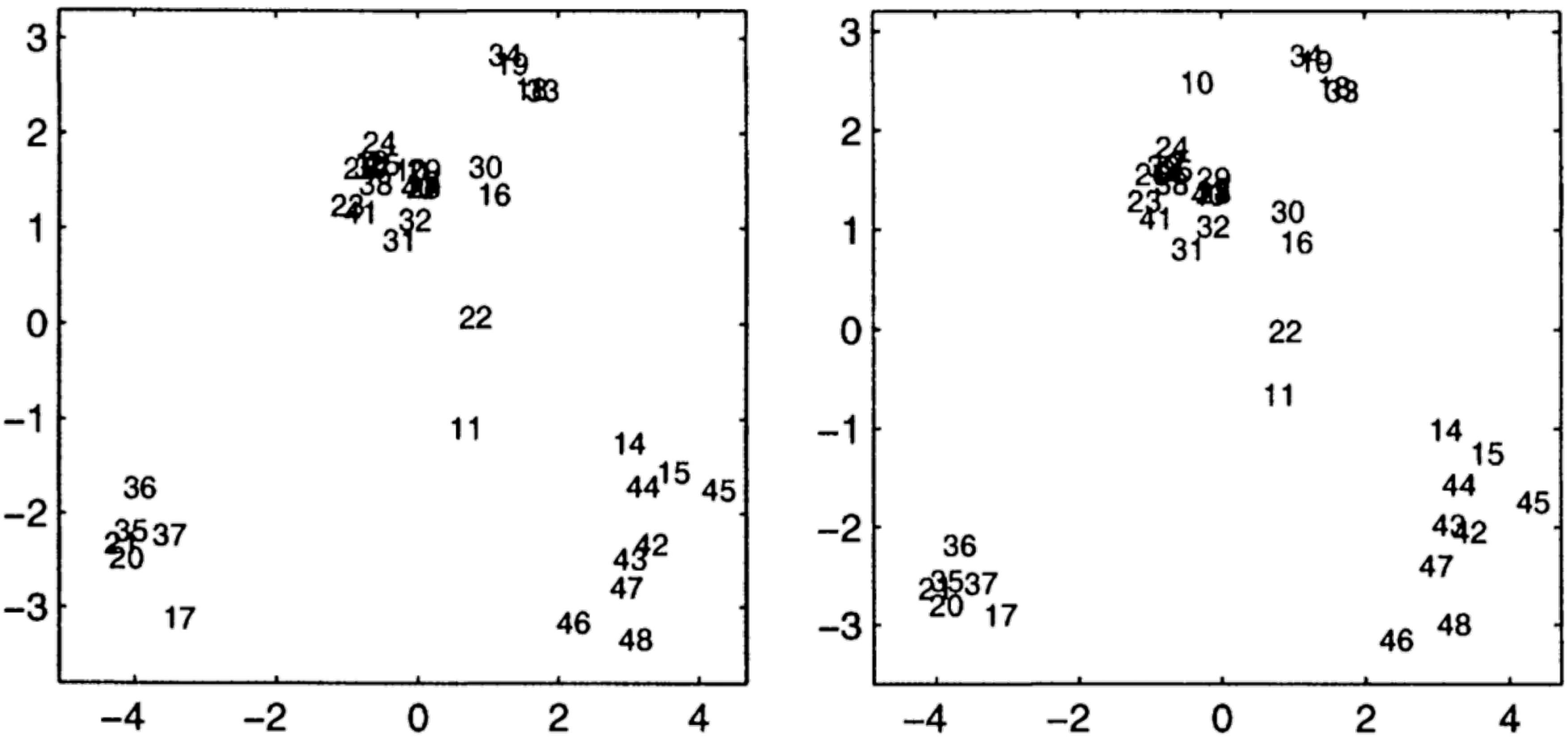


Principal Component Analysis



The idea in PCA is to identify variables that are correlated, such that we may as well represent the two as one variable.

Principal Component Analysis

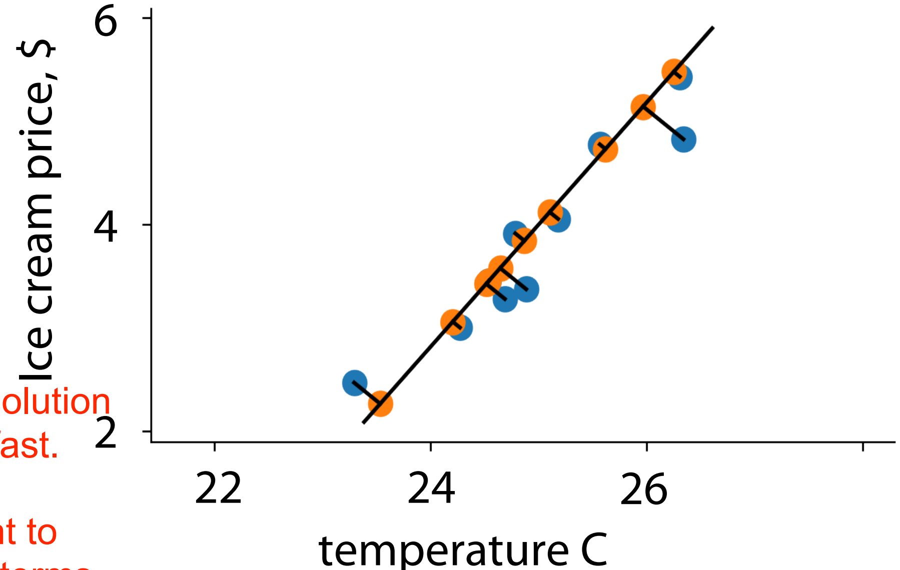


Projection of the Tobamovirus data by using PCA on the full data set and PPCA with 136 missing values

[source: Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis]

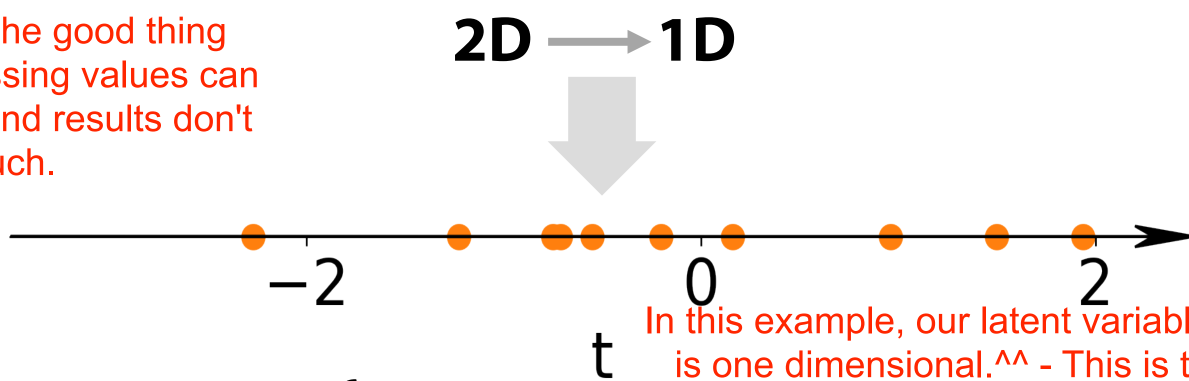
So the good thing about probabilistic PCA is that even with missing values, they are pretty much the same.

Principal Component Analysis



In PCA, there is an analytical solution to reduce dimensions. It is fast.

But sometimes, people want to formulate PCA in probabilistic terms. See previous slide. The good thing about PPCA is that missing values can still be accounted for and results don't vary as much.



The prior for the t_i is the standard normal. This means that the low dimensional projection will be somewhere around 0 and variance 1. (t_i follows the standard normal pdf)

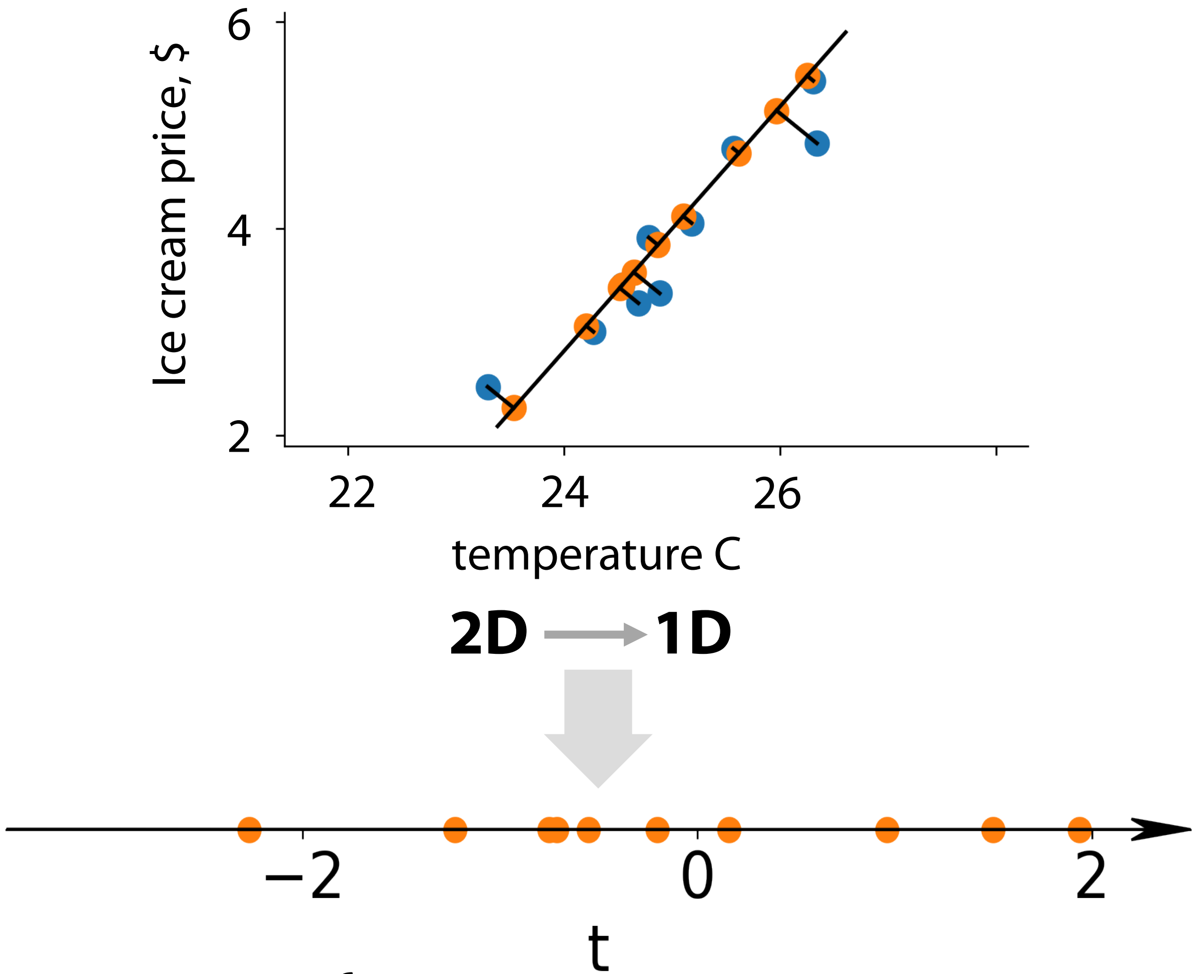
In this example, our latent variable t_i is one dimensional.^^ - This is the PCA'd data.

$$p(t_i) = \mathcal{N}(0, I)$$

$$x_i = (1, 1)t_i + (25, 4)$$

WE MAP our latent variable t_i into some two dimensional datapoint x_i. This translates to the full data.

Principal Component Analysis

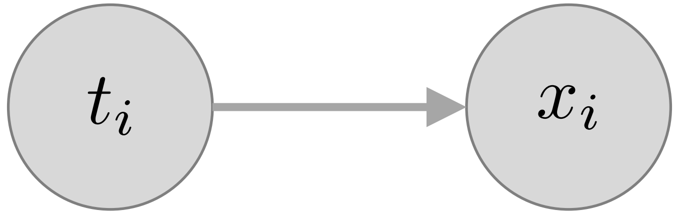


$$p(t_i) = \mathcal{N}(0, I)$$

$$x_i = W t_i + b + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \Sigma)$$

epsilon here is the error. Its some noise.

Principal Component Analysis



$$p(t_i) = \mathcal{N}(0, I)$$

$$p(x_i \mid t_i, \theta) = \mathcal{N}(Wt_i + b, \Sigma)$$

$$\max_{\theta} p(X \mid \theta) = \prod_{i=1}^{N} p(x_i \mid \theta)$$

N independent data points. So we do the product

^ this is the Maximum Likelihood.

^ this is the Maximum Likelihood.

$$= \prod_i \underbrace{\int p(x_i \mid t_i, \theta) p(t_i) dt_i}_{\text{conjugacy, } \mathcal{N}(\mu_i, \Sigma_i)}$$

We marginalize out t_i here^.
In general, this integral is intractable. So we use the EM algorithm to get an approximate.