Carnegie Mellon University
Heinz College

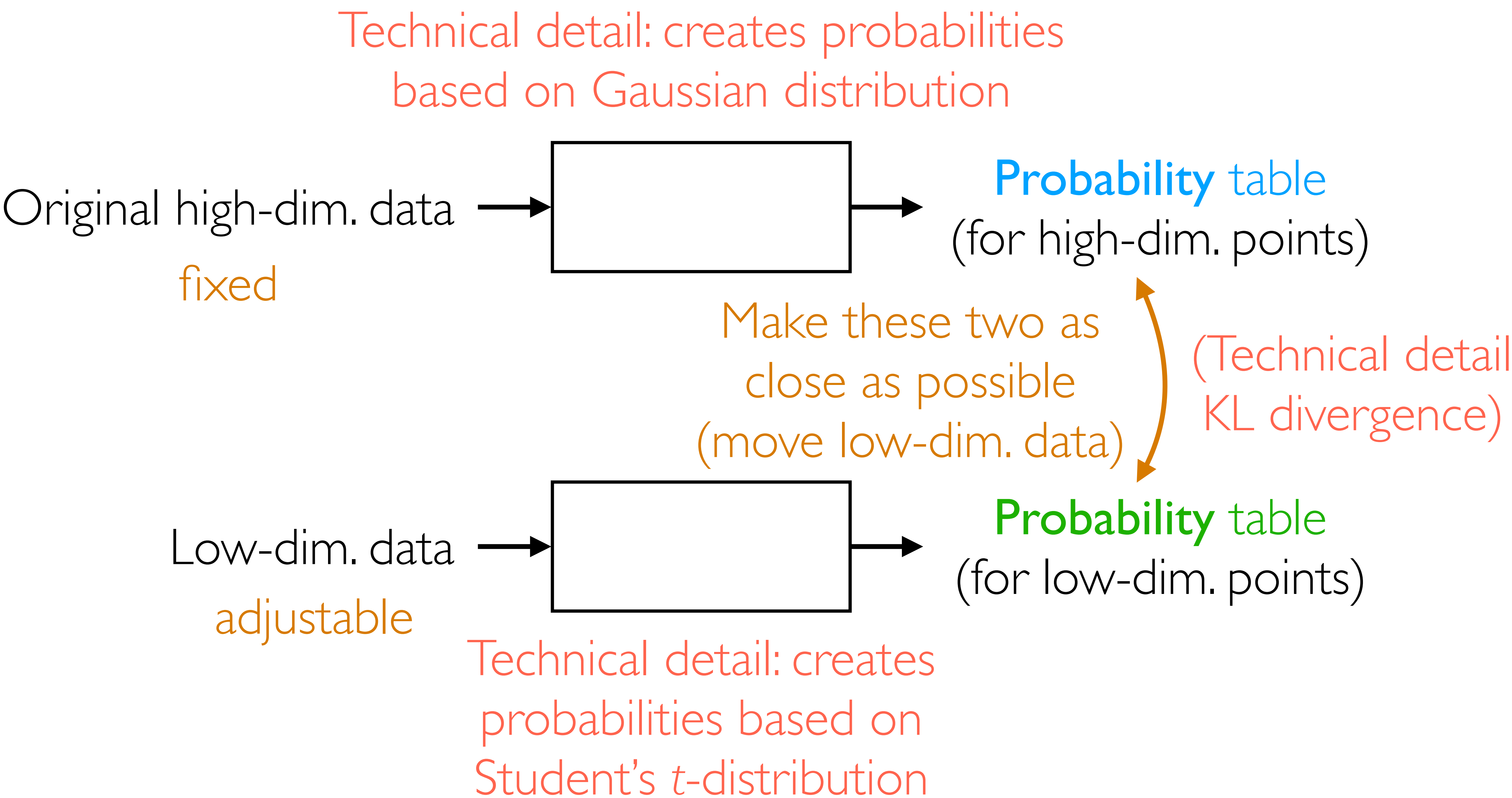
# Unstructured Data Analysis

## t-SNE: Some technical details

George Chen

For the purposes of this course, you do *not* need to know these technical details
(I'm providing these just to give you a flavor of what some algorithms are like)

# t-SNE




Technical details are in separate slides (posted on webpage)

# Technical Detail for t-SNE

## High-dimensional space: how to compute the probability table

Suppose there are $n$ high-dimensional points $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$

For a specific point $i$, point $i$ picks point $j$ ($\neq i$) to be a neighbor with probability:

$$p_{j|i} = \frac{\exp(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2})}{\sum_{k \neq i} \exp(-\frac{\|x_i - x_k\|^2}{2\sigma_i^2})}$$

$\boldsymbol{\sigma_i}$ (depends on $i$) controls the probability in which point $j$ would be picked by $i$ as a neighbor (think about when it gets close to 0 or when it explodes to $\boldsymbol{\infty}$)

$\boldsymbol{\sigma_i}$ is controlled by a knob called **perplexity**
(rough intuition: it is like the "number of nearest neighbors" in Isomap)

Points $i$ and $j$ are "similar" with probability: $$p_{i,j} = \frac{p_{j|i} + p_{i|j}}{2n}$$

This defines the blue distribution in the lecture slides

# Technical Detail for t-SNE

## Low-dimensional space: how to compute the probability table

Denote the $n$ low-dimensional points as $\mathbf{x_1}'$, $\mathbf{x_2}'$, …, $\mathbf{x_n}'$

Low-dim. points $i$ and $j$ are "similar" with probability: $$q_{i,j} = \frac{\frac{1}{1+\|x_i' - x_j'\|^2}}{\sum_{k \neq m} \frac{1}{1+\|x_k' - x_m'\|^2}}$$

This defines the green distribution in the lecture slides

## How to compare high/low-dimensional probability tables

Approximately minimize (with respect to $\mathbf{q_{i,j}}$) the following cost:

$$\sum_{i \neq j} p_{i,j} \log \frac{p_{i,j}}{q_{i,j}}$$

This cost is called the "KL divergence" between distributions $\boldsymbol{p}$ and $\boldsymbol{q}$