# Implicit Maximum Likelihood Estimation

Ke Li
Jitendra Malik
For work on super-resolution, also joint with Shichong Peng
For work on image synthesis from scene layouts, also joint with Tianhao Zhang

Berkeley
UNIVERSITY OF CALIFORNIA

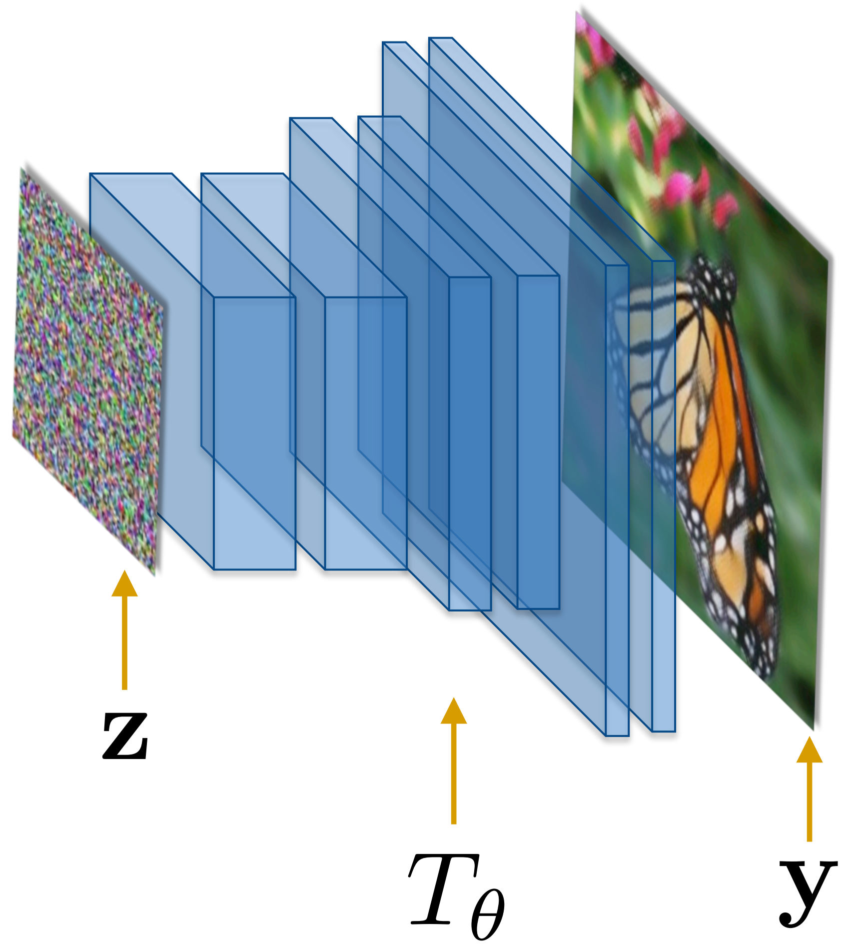
# Probabilistic Models

- Prescribed Probabilistic Models:
  - Models whose densities can be expressed explicitly.
  - E.g.: Models in the exponential family
- Implicit Probabilistic Models:
  - Models defined in terms of a sampling procedure.
  - E.g.:
    1. Sample $\mathbf{z} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}\right)$
    2. Return $\mathbf{y} = T_\theta(\mathbf{z})$
- Implicit probabilistic models offer more modelling flexibility, since $T_\theta$ can be an arbitrary function, e.g. neural nets.

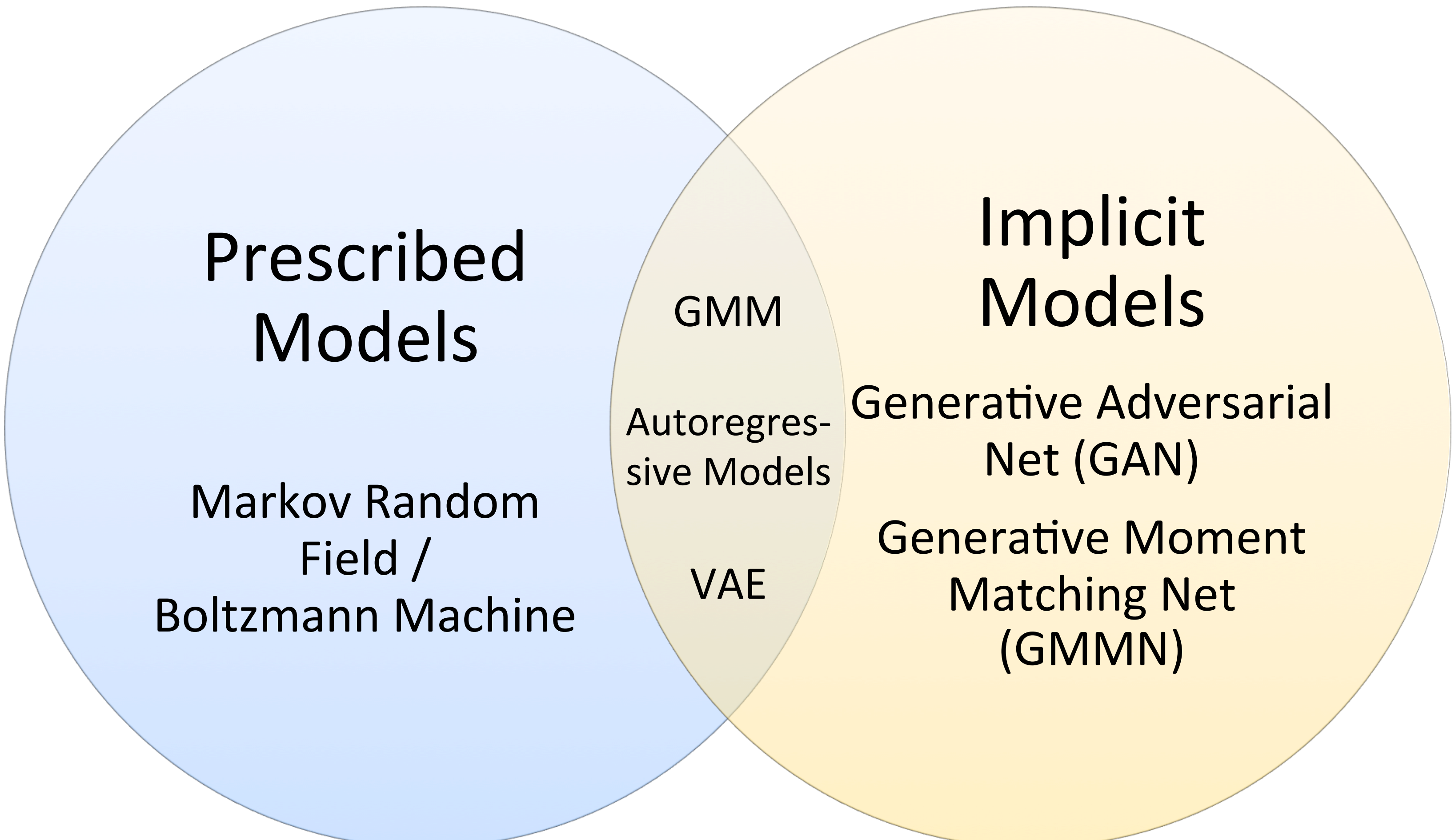
Berkeley
UNIVERSITY OF CALIFORNIA

# Implicit Probabilistic Model

$$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$
$$\mathbf{y} = T_\theta(\mathbf{z})$$



# Probabilistic Models



# Parameter Estimation

- For prescribed probabilistic models, a variety of methods:
  - If density is tractable: maximum likelihood
  - If density is intractable: maximize some approximation to likelihood, e.g.: variational Bayes, contrastive divergence, pseudolikelihood
- For implicit probabilistic models, the form of the likelihood is unknown.
  - Cannot maximize likelihood.

# Agenda

1. *Implicit Maximum Likelihood Estimation*
2. Comparison to Generative Adversarial Nets (GANs)
3. Why Maximum Likelihood
4. Equivalence to Maximum Likelihood
5. Fast Nearest Neighbour Search
6. Applications to Conditional Image Synthesis

# Implicit Maximum Likelihood Estimation (IMLE)

# Motivation

- How do we train implicit probabilistic models?
  - By definition, we don't have access to likelihood, but have access to samples.
- This problem is known as likelihood-free inference – we need to infer parameters without any access to likelihood or derived quantities.
- Various methods have been proposed:
  - Unfortunately, they do not maximize likelihood.
  - In fact, they have known biases, both in theory and in practice.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

Berkeley
UNIVERSITY OF CALIFORNIA

# Motivation

- How do we train implicit probabilistic models?
  - By definition, we don't have access to likelihood, but have access to samples.
- Th… …e –
  w… …o
  lik…
- Va…

Can we maximize likelihood without computing likelihood?

  - Unfortunately, they do not maximize likelihood.
  - In fact, they have known biases, both in theory and in practice.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Motivation

- A model distribution that maximizes likelihood should have high density at each data example.
- So, samples from the model would be more likely to lie near data examples than elsewhere.
- Intuitively, to maximize likelihood, we can simply encourage this to happen.

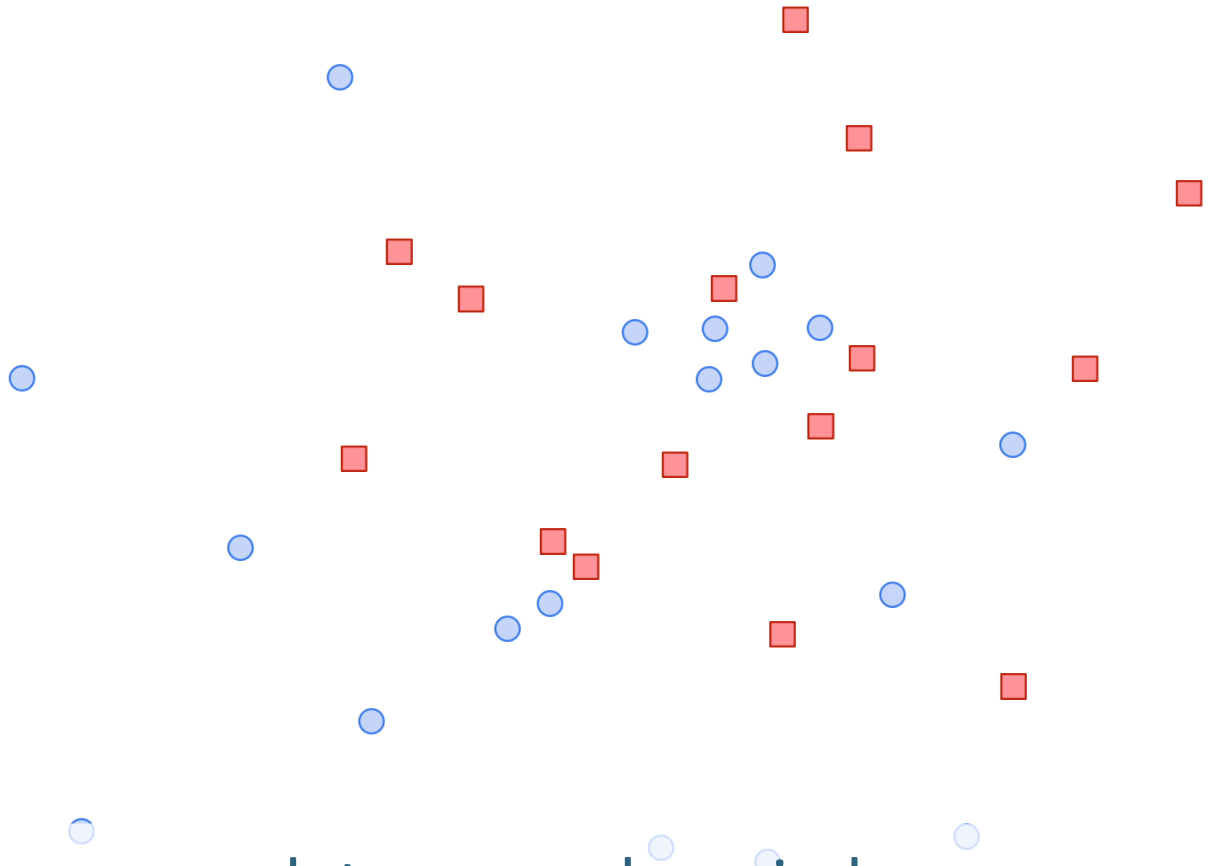
K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Method

- Simple idea:
  1. Generate a batch of i.i.d. samples.
  2. Find the nearest sample to each data example.
  3. Adjust the parameters so that the nearest sample is pulled towards each data example.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

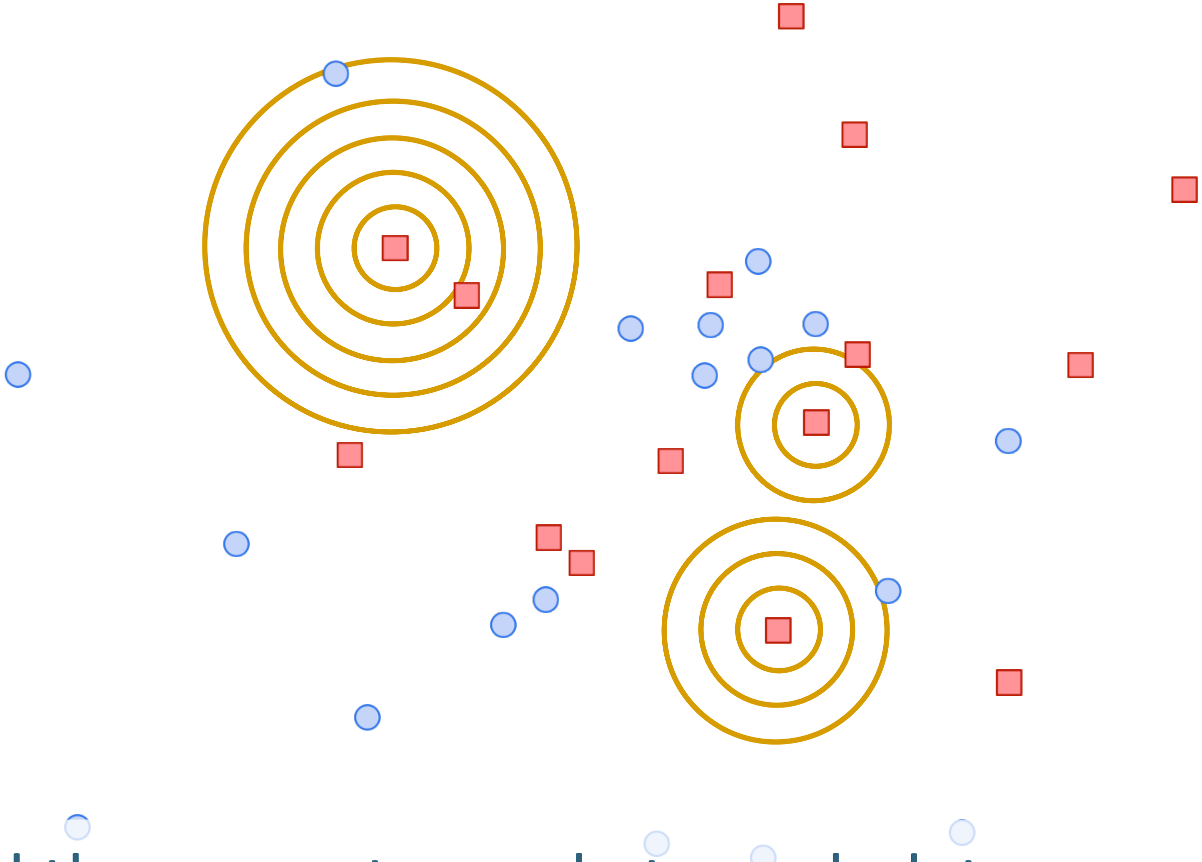
# Method



Squares are data examples; circles are samples.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

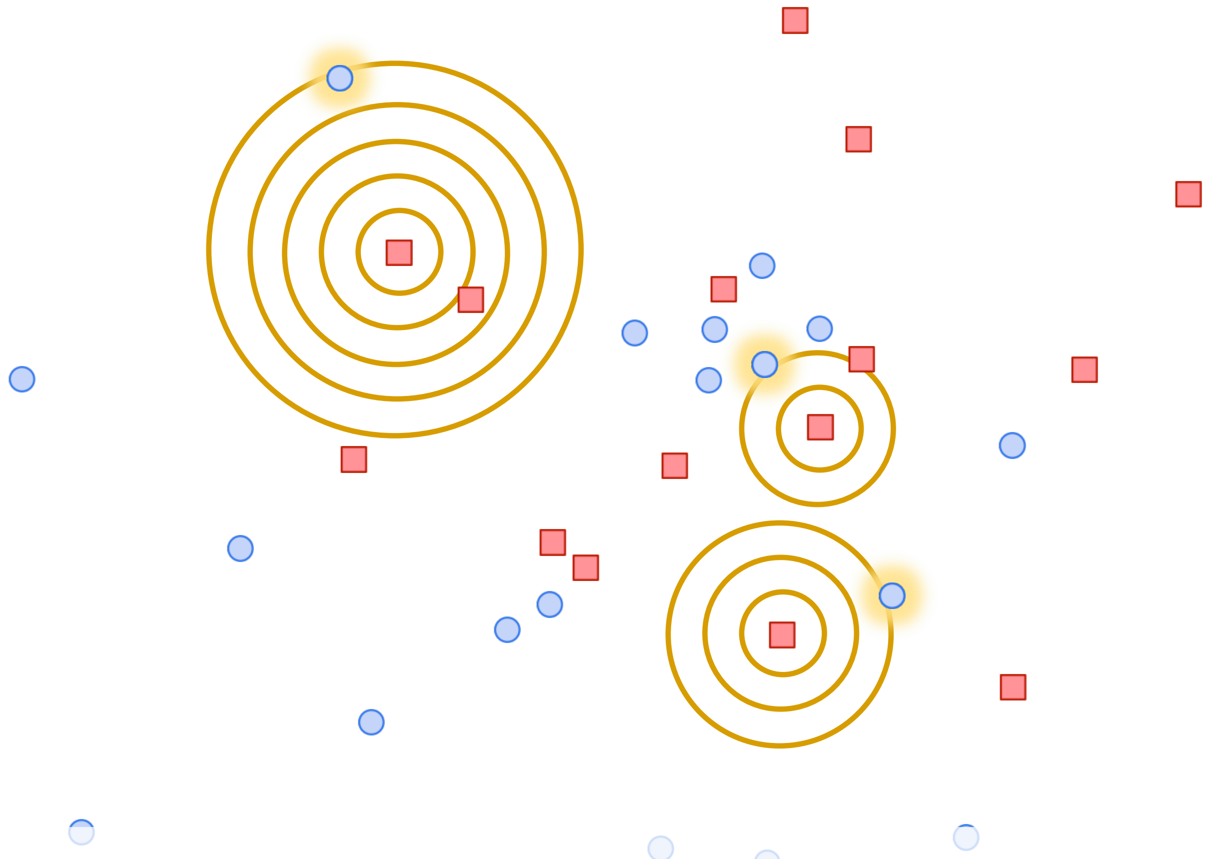
# Method



Find the nearest sample to each data example.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Method

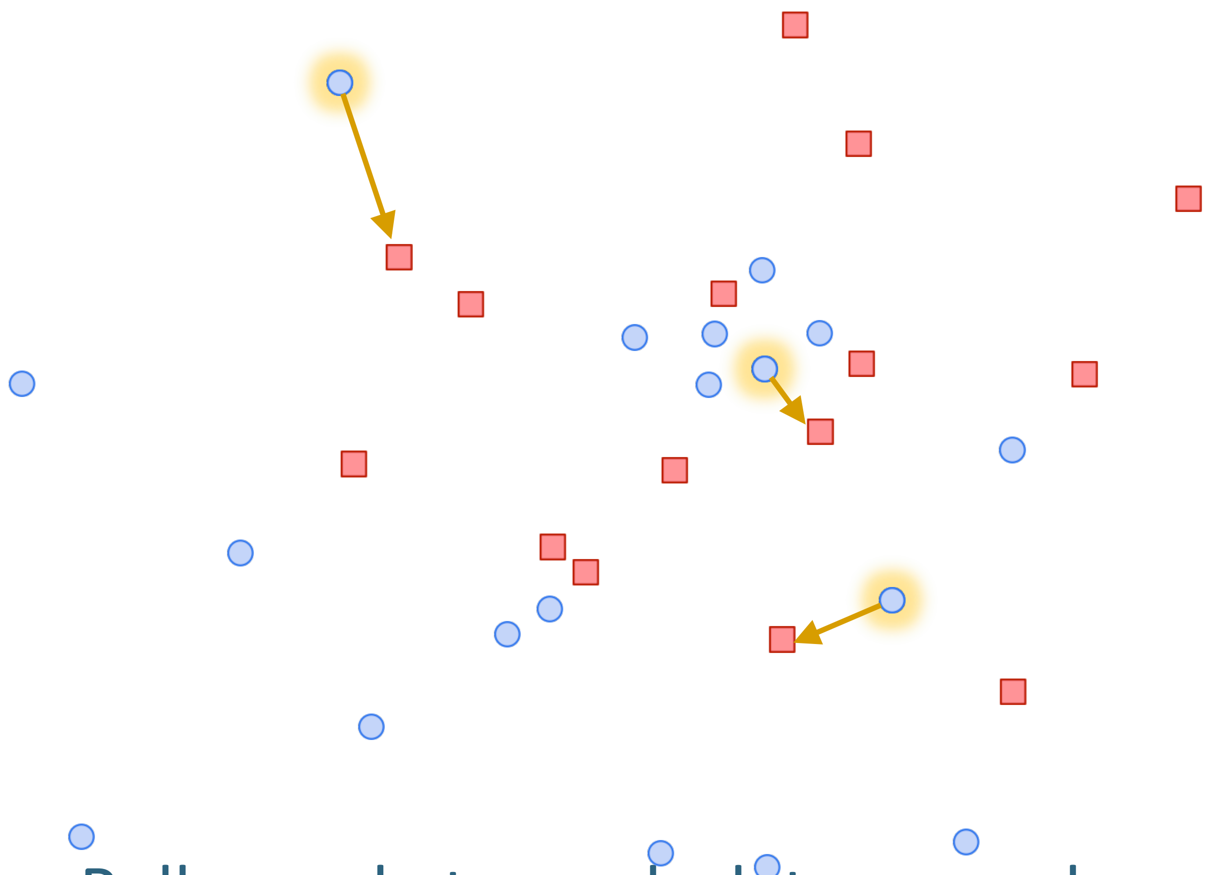


Nearest samples are found.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

Berkeley
UNIVERSITY OF CALIFORNIA

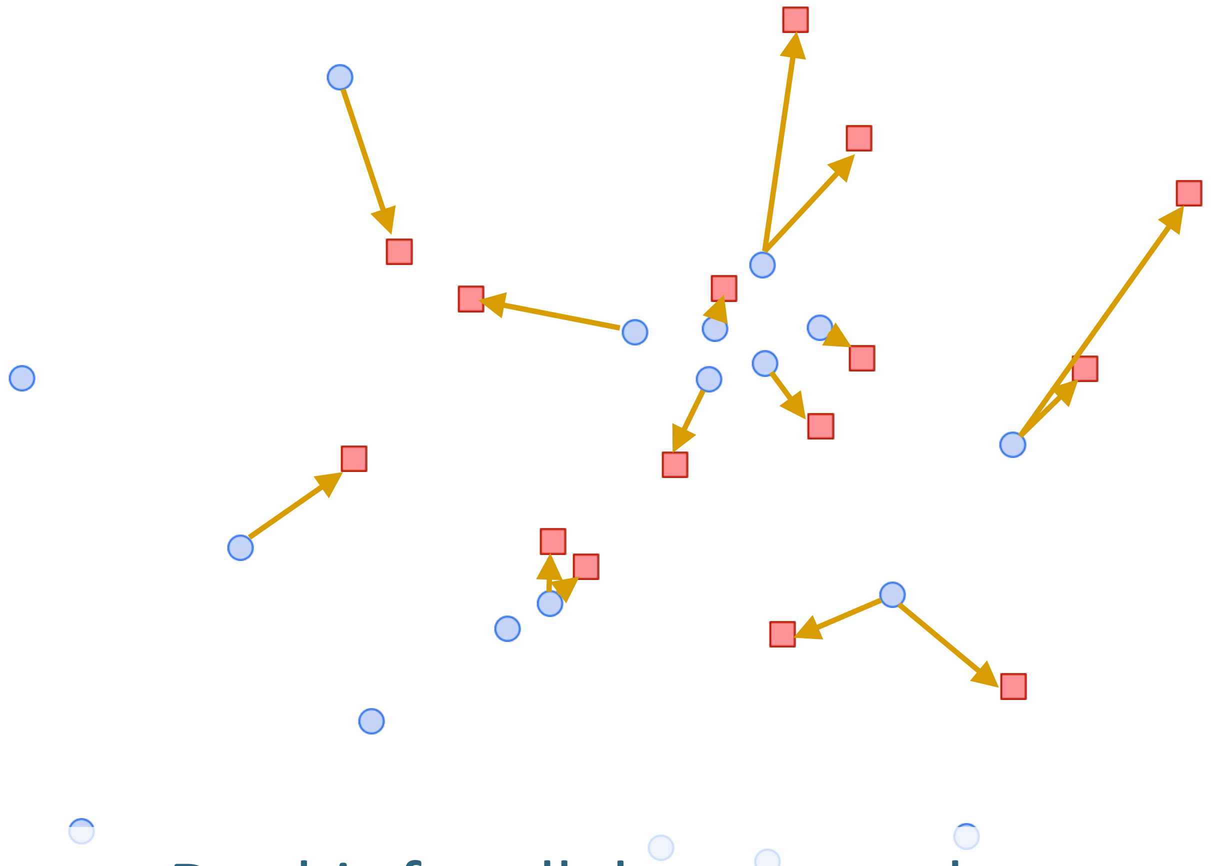
# Method



Pull sample towards data example.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Method



Do this for all data examples.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Method

- The Implicit Maximum Likelihood Estimator is defined as:

$\mathbf{x}_1, \dots, \mathbf{x}_n$: data examples
$\tilde{\mathbf{x}}_1^\theta, \dots, \tilde{\mathbf{x}}_m^\theta$: i.i.d. samples from $P_\theta$

$$\hat{\theta}_{\text{IMLE}} := \arg\min_{\theta} \mathbb{E}_{\tilde{\mathbf{x}}_1^\theta, \dots, \tilde{\mathbf{x}}_m^\theta} \left[ \sum_{i=1}^{n} \min_{j \in [m]} \left\| \tilde{\mathbf{x}}_j^\theta - \mathbf{x}_i \right\|_2^2 \right]$$

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Method

- The Implicit M Nearest Neighbour Search or is defined as:

*Nearest Neighbour Search*: Used to be hard in high dimensions due to the curse of dimensionality. Will discuss how to overcome this later.

$\mathbf{x}_1, \ldots, \mathbf{x}_n$

$\tilde{\mathbf{x}}_1^\theta, \ldots, \tilde{\mathbf{x}}_m^\theta$: i.i.d. samples fr $\theta$

$$\hat{\theta}_{\mathrm{IMLE}} := \arg\min_\theta \mathbb{E}_{\tilde{\mathbf{x}}_1^\theta, \ldots, \tilde{\mathbf{x}}_m^\theta}\left[\sum_{i=1}^{n} \boxed{\min_{j \in [m]} \left\|\tilde{\mathbf{x}}_j^\theta - \mathbf{x}_i\right\|_2^2}\right]$$

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Method

**Algorithm 1** Implicit maximum likelihood estimation (IMLE) procedure

**Require:** The dataset $D = \{\mathbf{x}_i\}_{i=1}^{n}$ and a sampling mechanism for the implicit model $P_\theta$

Initialize $\theta$ to a random vector

**for** $k = 1$ **to** $K$ **do**

Draw i.i.d. samples $\tilde{\mathbf{x}}_1^\theta, \dots, \tilde{\mathbf{x}}_m^\theta$ from $P_\theta$

Pick a random batch $S \subseteq \{1, \dots, n\}$

$\sigma(i) \leftarrow \arg\min_j \left\| \mathbf{x}_i - \tilde{\mathbf{x}}_j^\theta \right\|_2^2 \;\; \forall i \in S$

**for** $l = 1$ **to** $L$ **do**

Pick a random mini-batch $\tilde{S} \subseteq S$

$\theta \leftarrow \theta - \eta \nabla_\theta \left( \frac{n}{|\tilde{S}|} \sum_{i \in \tilde{S}} \left\| \mathbf{x}_i - \tilde{\mathbf{x}}_{\sigma(i)}^\theta \right\|_2^2 \right)$

**end for**

**end for**

**return** $\theta$

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley UNIVERSITY OF CALIFORNIA

# Overview

1. Implicit Maximum Likelihood Estimation
2. *Comparison to Generative Adversarial Nets (GANs)*
3. Why Maximum Likelihood
4. Equivalence to Maximum Likelihood
5. Fast Nearest Neighbour Search
6. Applications to Conditional Image Synthesis

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Comparison to Generative Adversarial Nets (GANs)

# Generative Adversarial Nets

- Generative Adversarial Nets (Goodfellow et al., 2014; Gutmann et al., 2014) is perhaps the most popular method for training implicit probabilistic models.
  - Key Idea: Introduce a classifier, called a discriminator, and train it to differentiate between model samples and the data.
  - Train the generative model ("generator") to fool the discriminator.

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{\mathbf{x}\sim p_{data}} \left[\log D_{\theta_D}(\mathbf{x})\right] + \mathbb{E}_{\mathbf{z}\sim p_z} \left[\log \left(1 - D_{\theta_D}\left(G_{\theta_G}(\mathbf{z})\right)\right)\right]$$

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

Berkeley
UNIVERSITY OF CALIFORNIA

# Generative Adversarial Nets

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{\mathbf{x}\sim p_{data}} \left[\log D_{\theta_D}(\mathbf{x})\right] + \mathbb{E}_{\mathbf{z}\sim p_z} \left[\log \left(1 - D_{\theta_D}\left(G_{\theta_G}(\mathbf{z})\right)\right)\right]$$

- Alternating gradient ascent-descent is used to optimize the objective, which performs the following updates iteratively:

$$\theta_D \leftarrow \theta_D + \gamma_D \nabla_{\theta_D} \left( \frac{1}{m} \sum_{i=1}^{m} \log D_{\theta_D}(\mathbf{x}^{(i)}) + \frac{1}{n} \sum_{j=1}^{n} \log \left(1 - D_{\theta_D}\left(G_{\theta_G}(\mathbf{z}^{(j)})\right)\right) \right)$$

$$\theta_G \leftarrow \theta_G - \gamma_G \nabla_{\theta_G} \left( \frac{1}{n} \sum_{j=1}^{n} \log \left(1 - D_{\theta_D}\left(G_{\theta_G}(\mathbf{z}^{(j)})\right)\right) \right)$$

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Generative Adversarial Nets

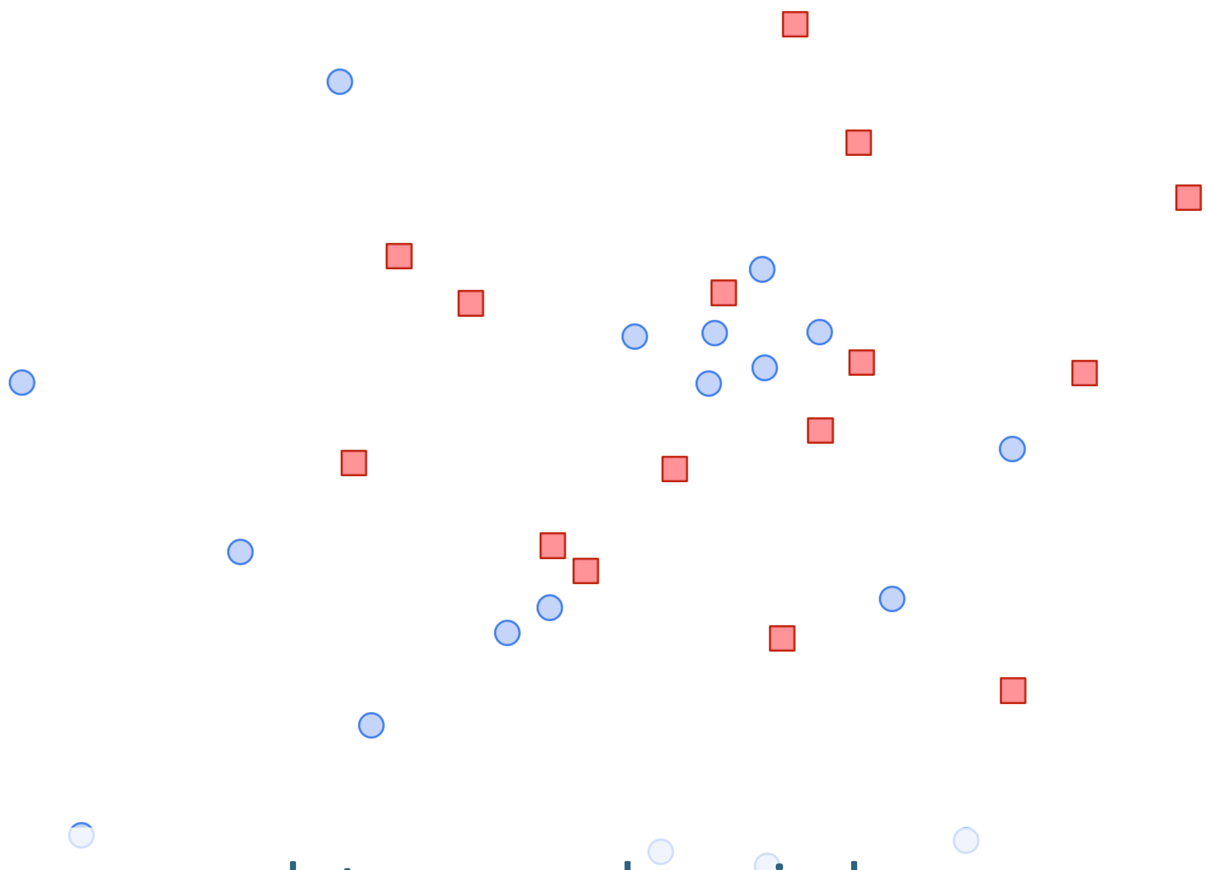
- Consider the case where the discriminator is a 1-nearest neighbour classifier.
- Would this be the same as IMLE?

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

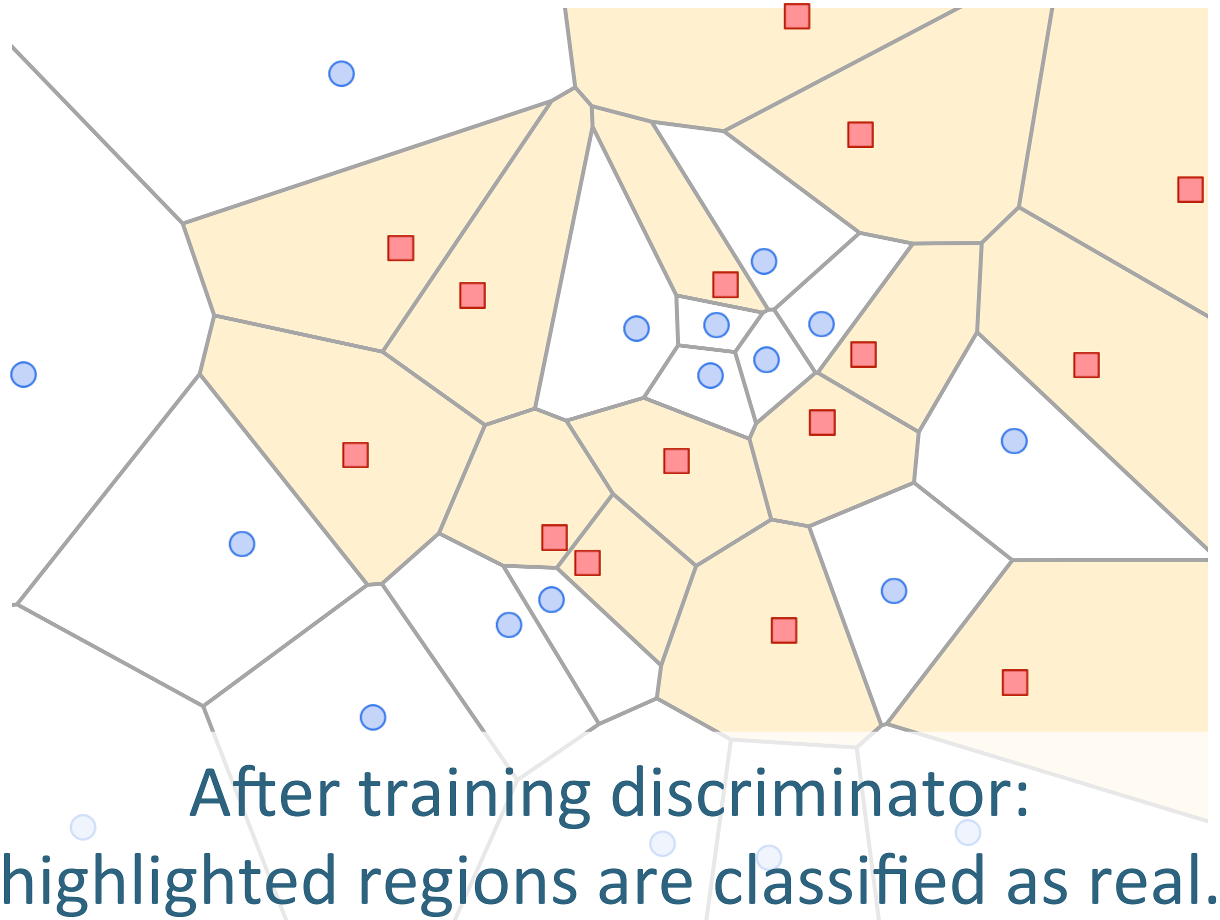
# Generative Adversarial Nets



Squares are data examples; circles are samples.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Generative Adversarial Nets



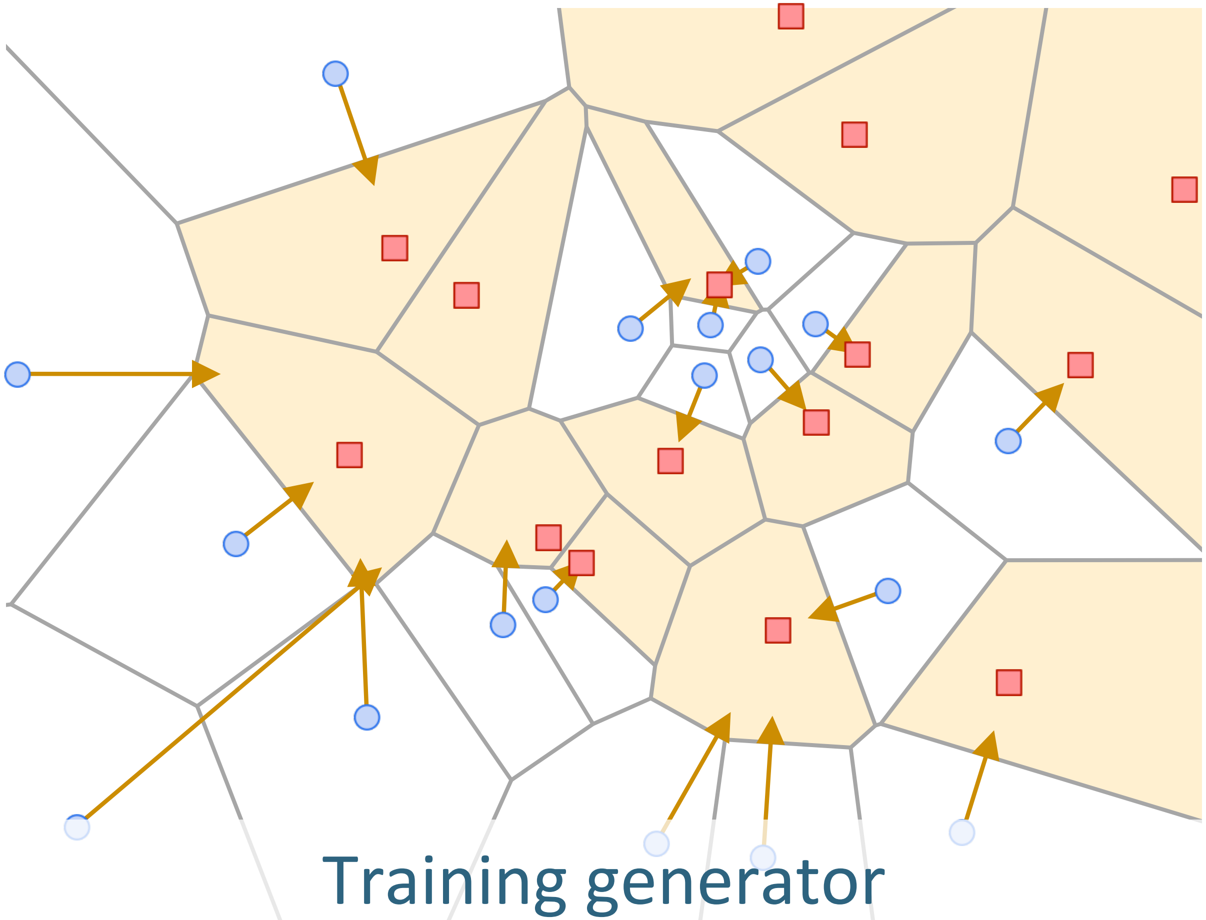
After training discriminator:
highlighted regions are classified as real.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Generative Adversarial Nets

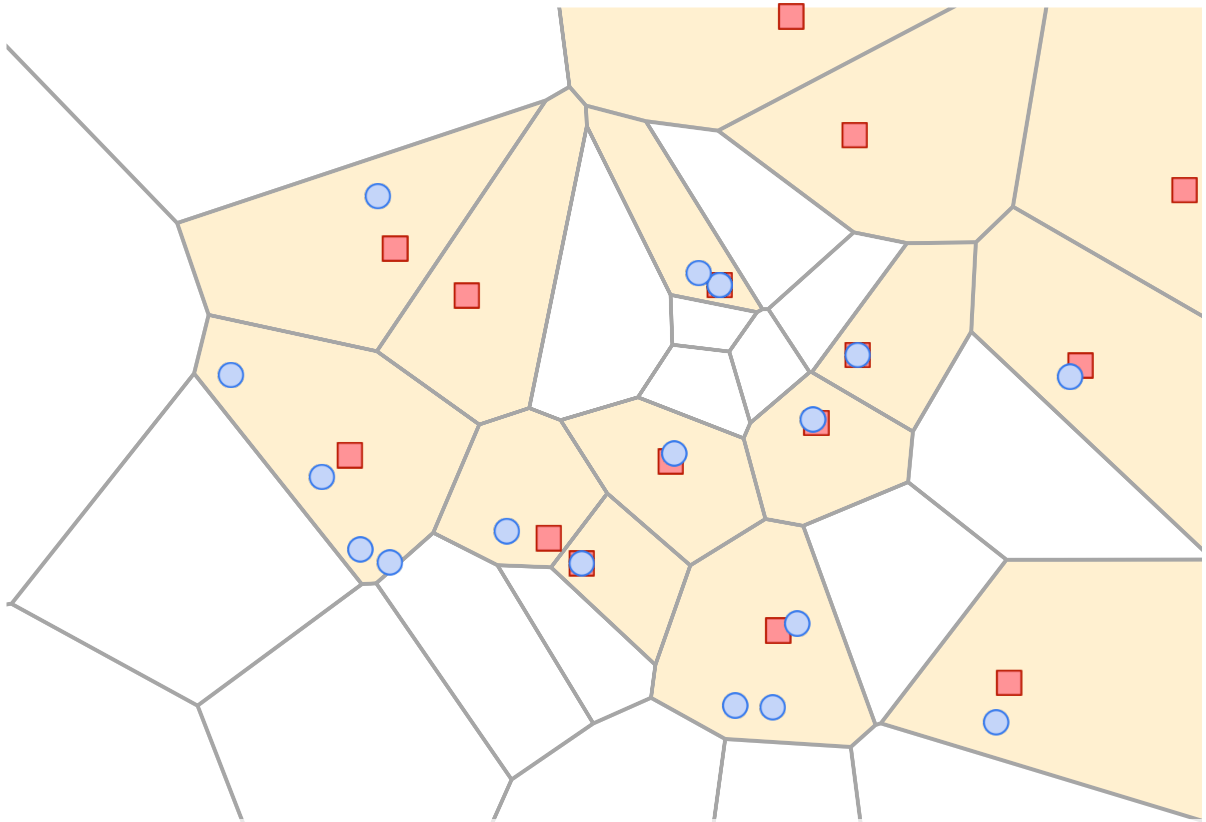


Training generator

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

Berkeley
UNIVERSITY OF CALIFORNIA

# Generative Adversarial Nets



After training generator

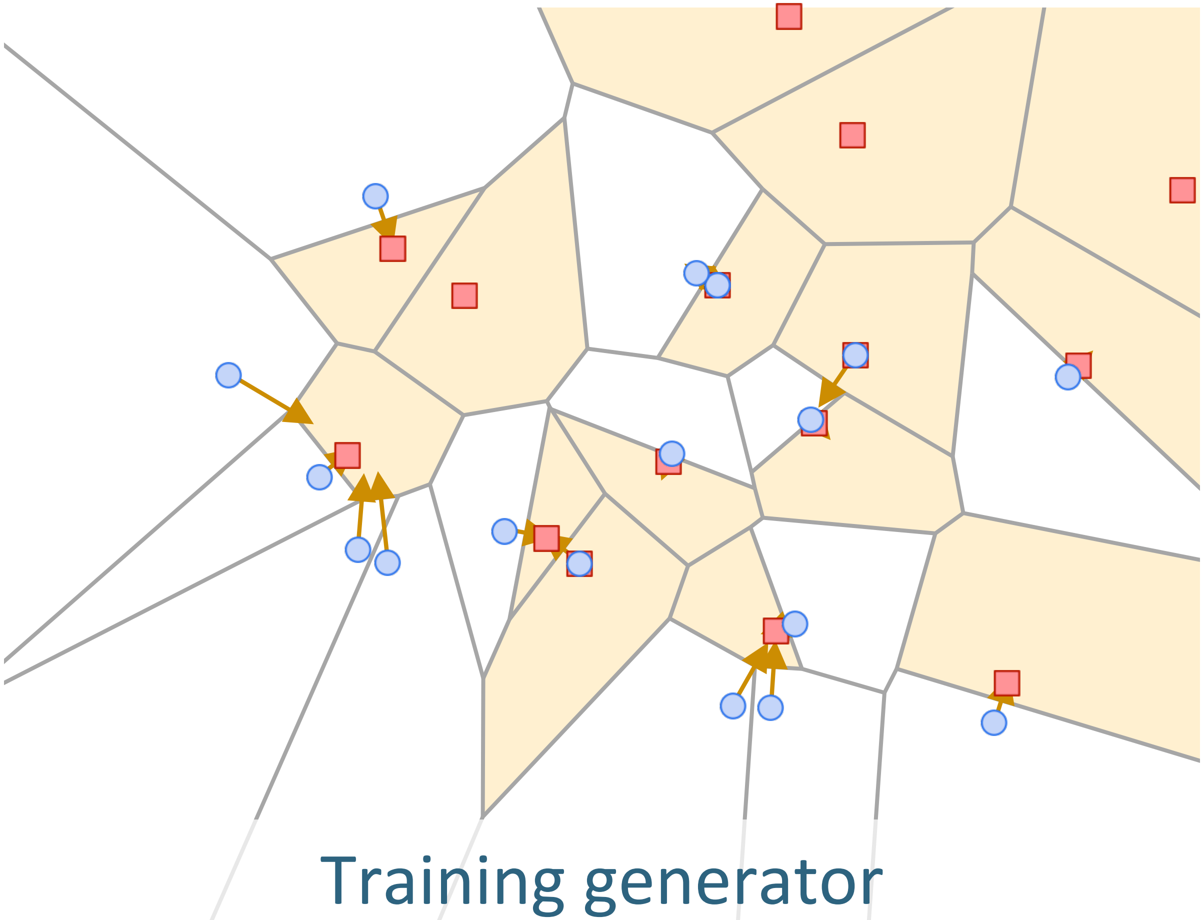
K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Generative Adversarial Nets

After training discriminator

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Generative Adversarial Nets



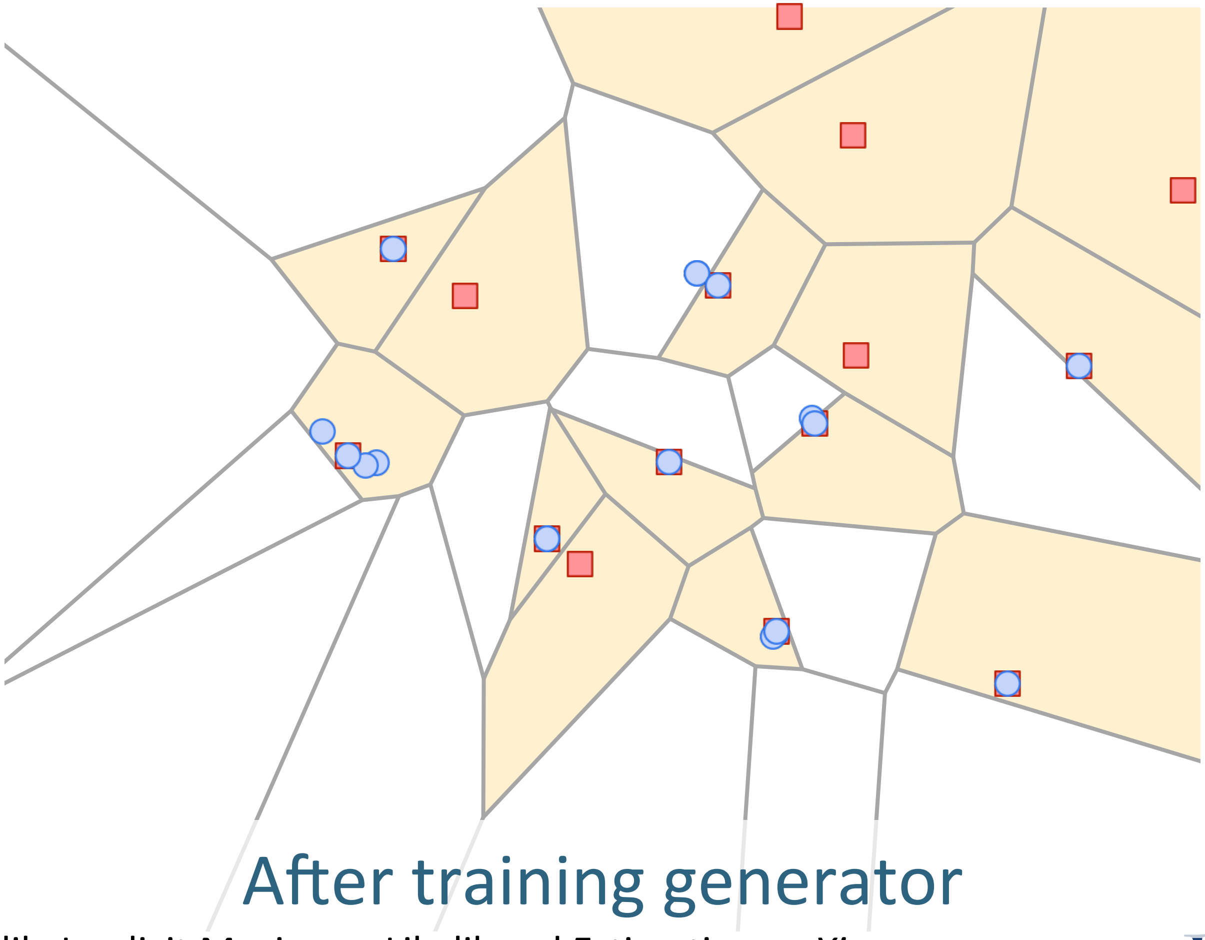
Training generator

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Generative Adversarial Nets



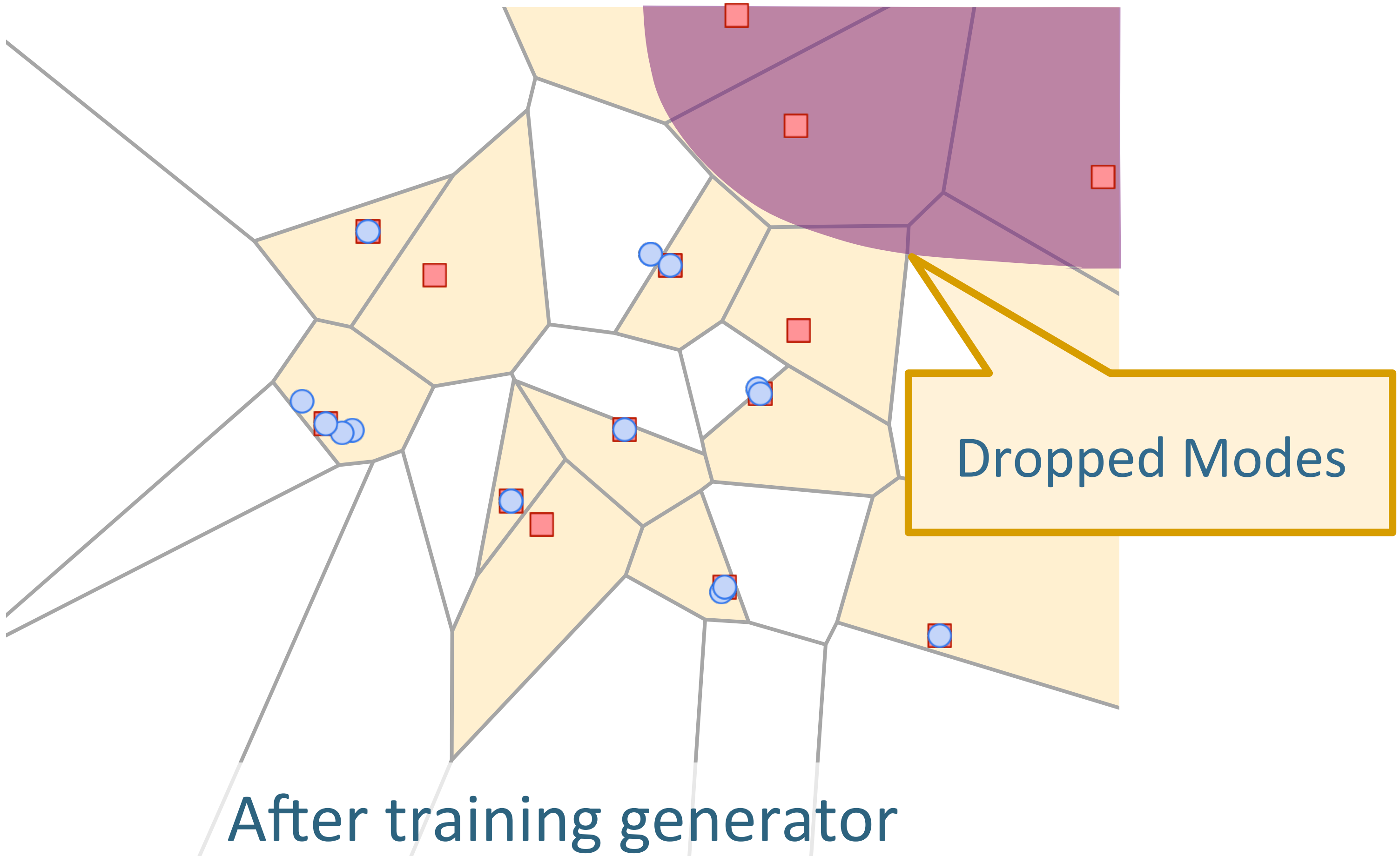
After training generator

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Generative Adversarial Nets



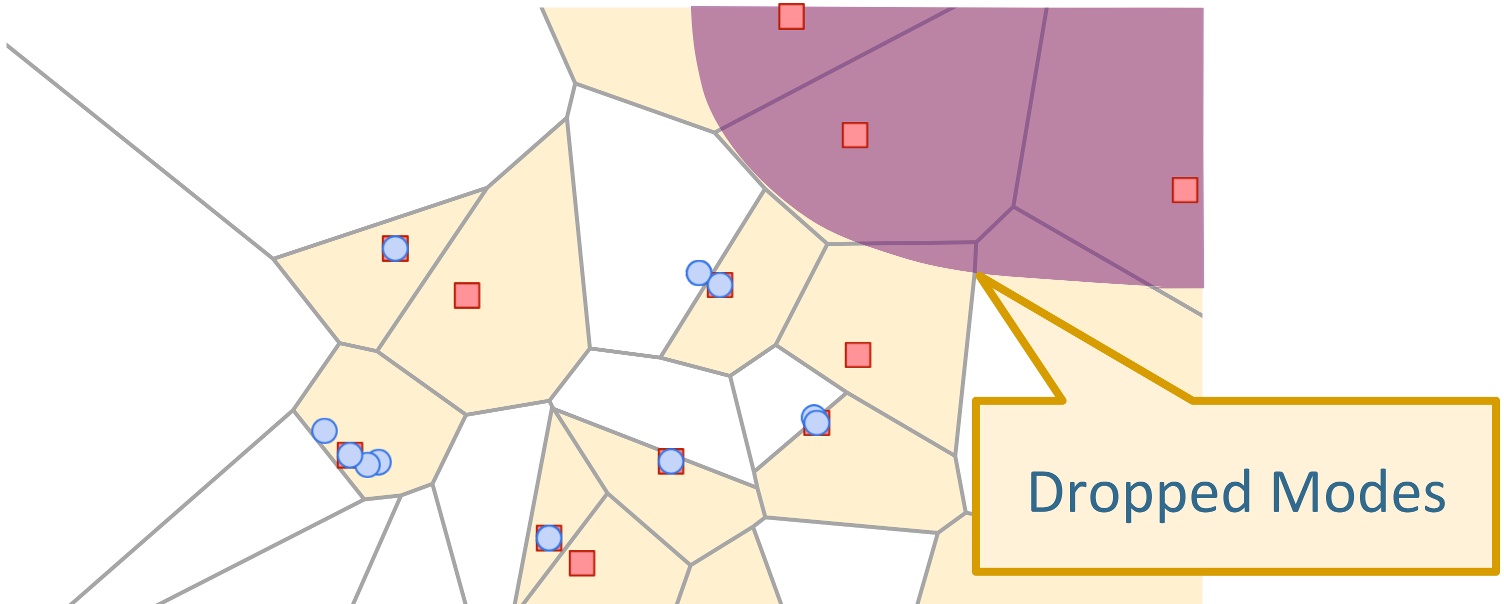

After training generator

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

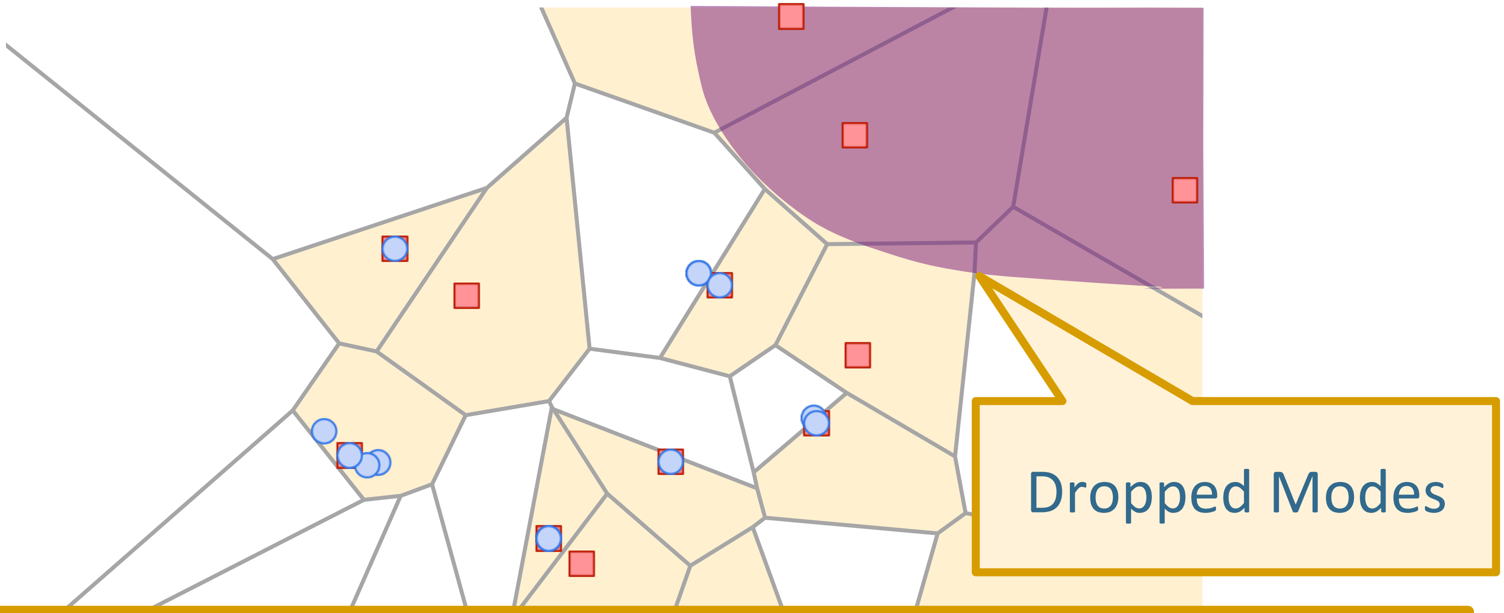
# Generative Adversarial Nets




GANs only care about making each *sample* similar to some *data example*; it does not care about whether each *data example* is similar to some *sample*.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Generative Adversarial Nets



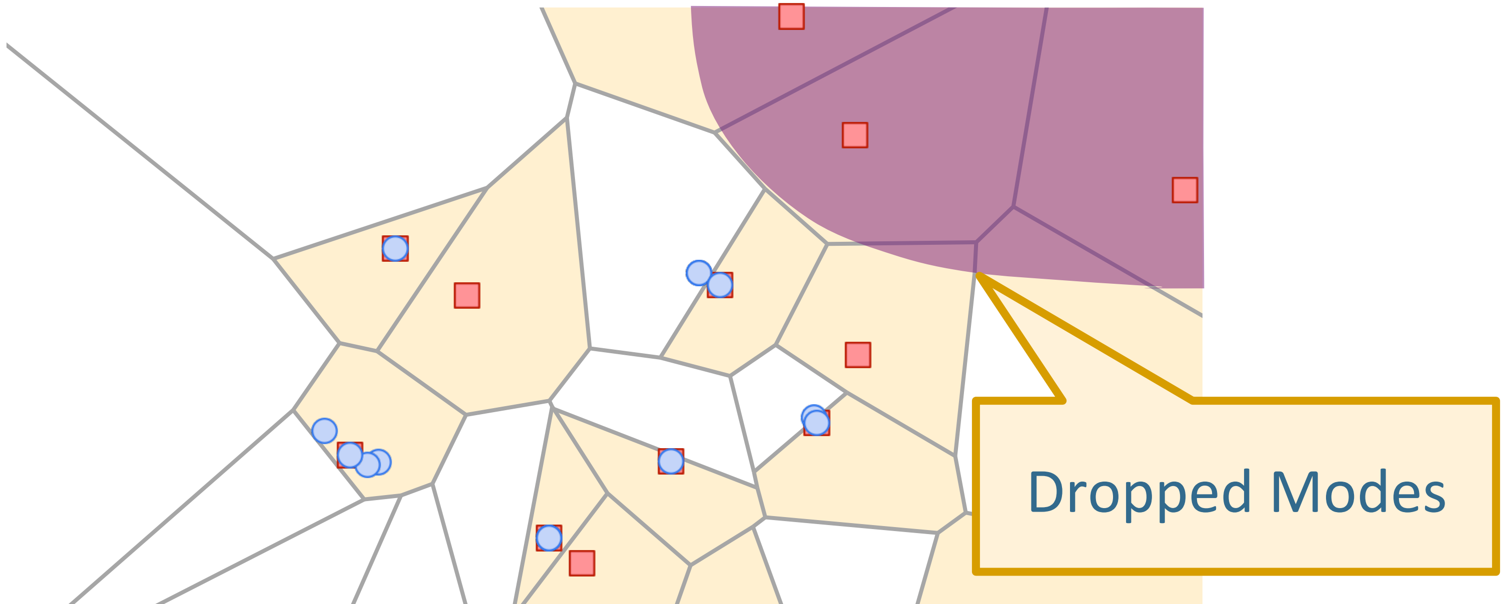

Each sample is essentially pushed towards the nearest data example. This is the opposite of IMLE.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley UNIVERSITY OF CALIFORNIA

# Generative Adversarial Nets




The direction in which nearest neighbour search is performed matters.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

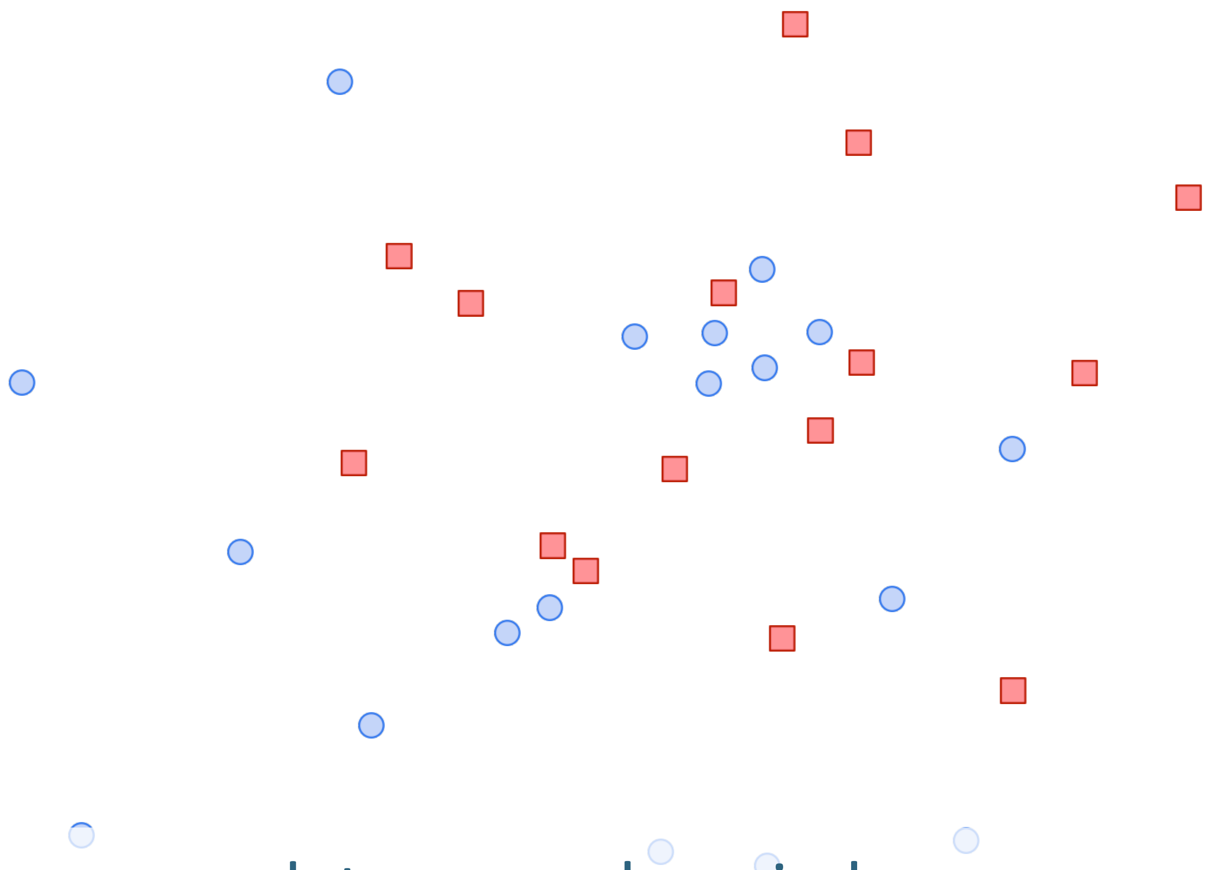
Berkeley
UNIVERSITY OF CALIFORNIA

# Maximize Likelihood of Samples Under Kernel Density Estimate

- What if we estimate the data distribution using a kernel density estimator (KDE) and then maximize the likelihood of samples under the estimated data distribution?
- In other words, what if the discriminator is a KDE?

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

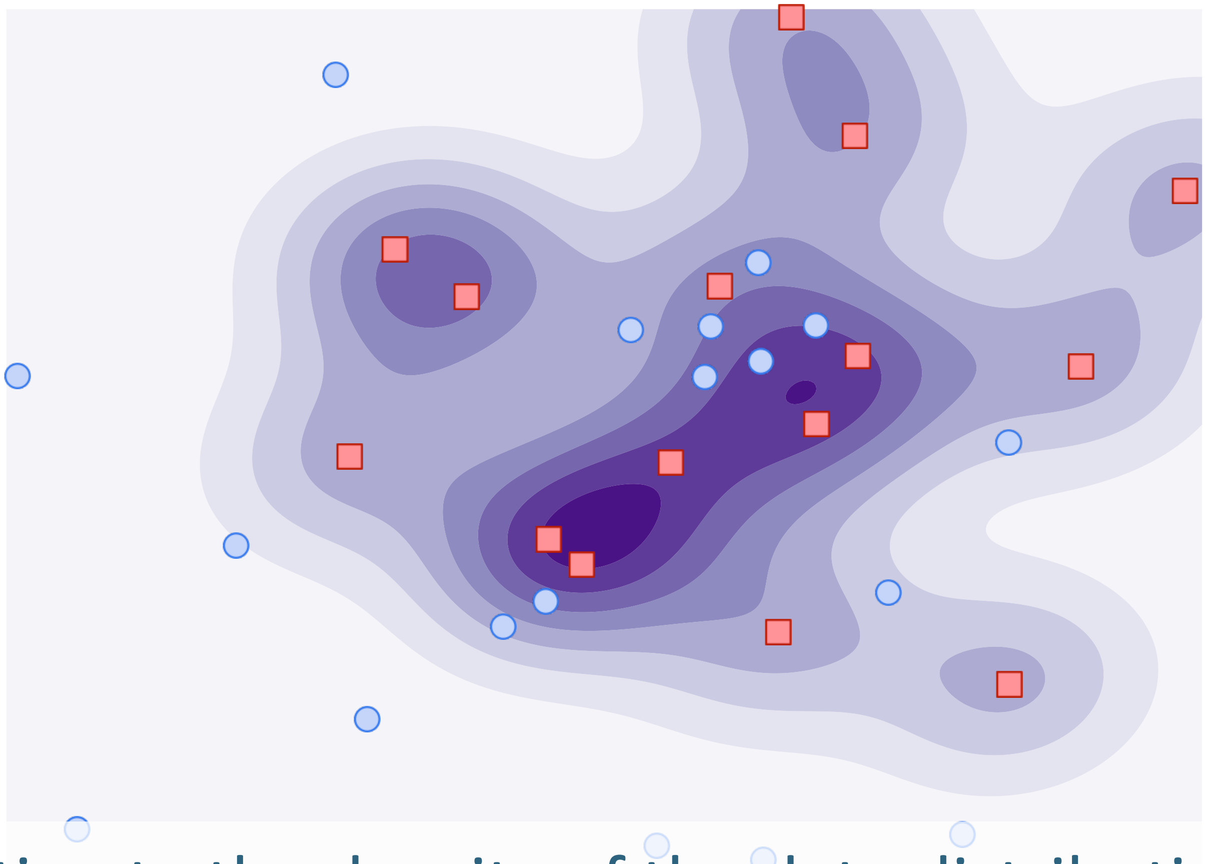
# Maximize Likelihood of Samples Under Kernel Density Estimate



Squares are data examples; circles are samples.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

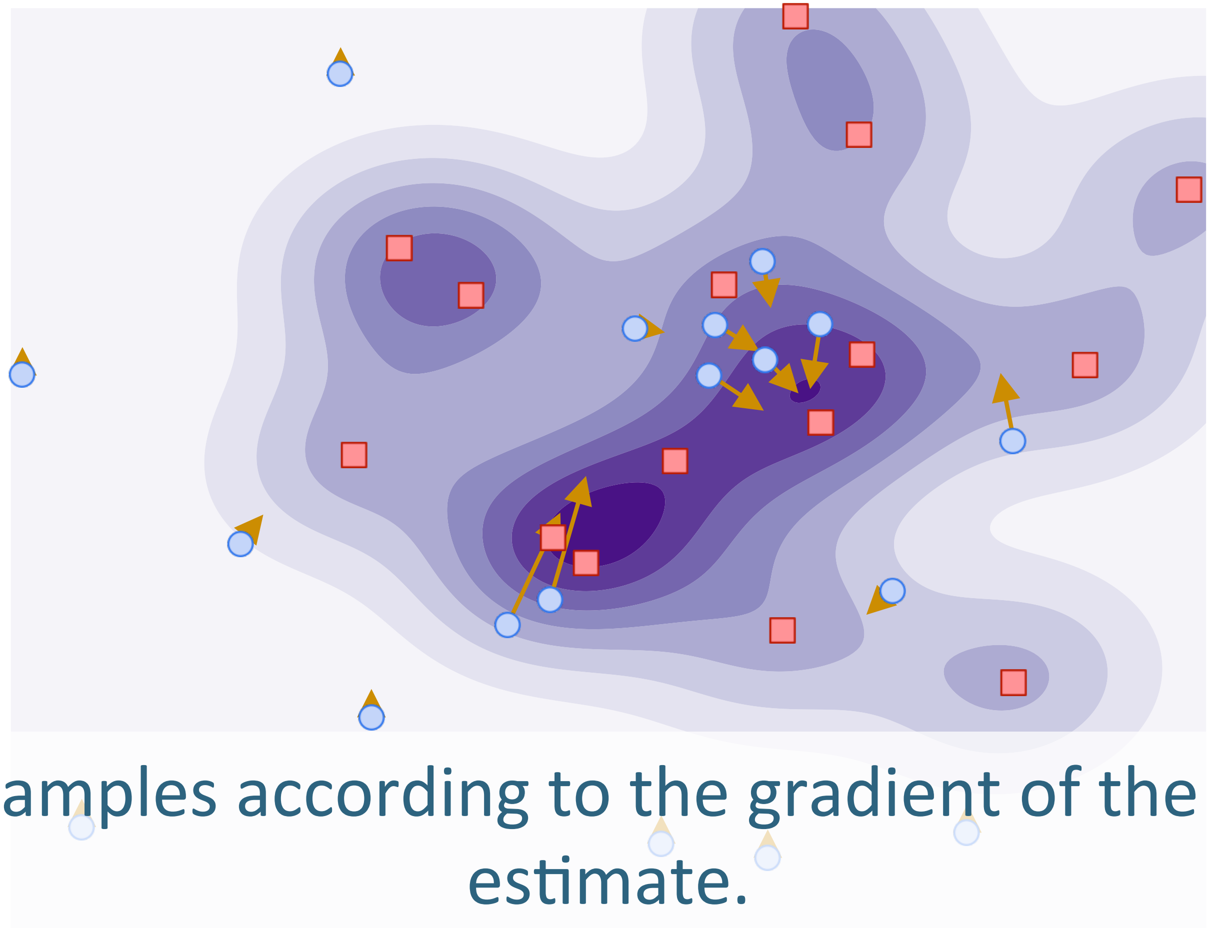
# Maximize Likelihood of Samples Under Kernel Density Estimate



Estimate the density of the data distribution.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Maximize Likelihood of Samples Under Kernel Density Estimate



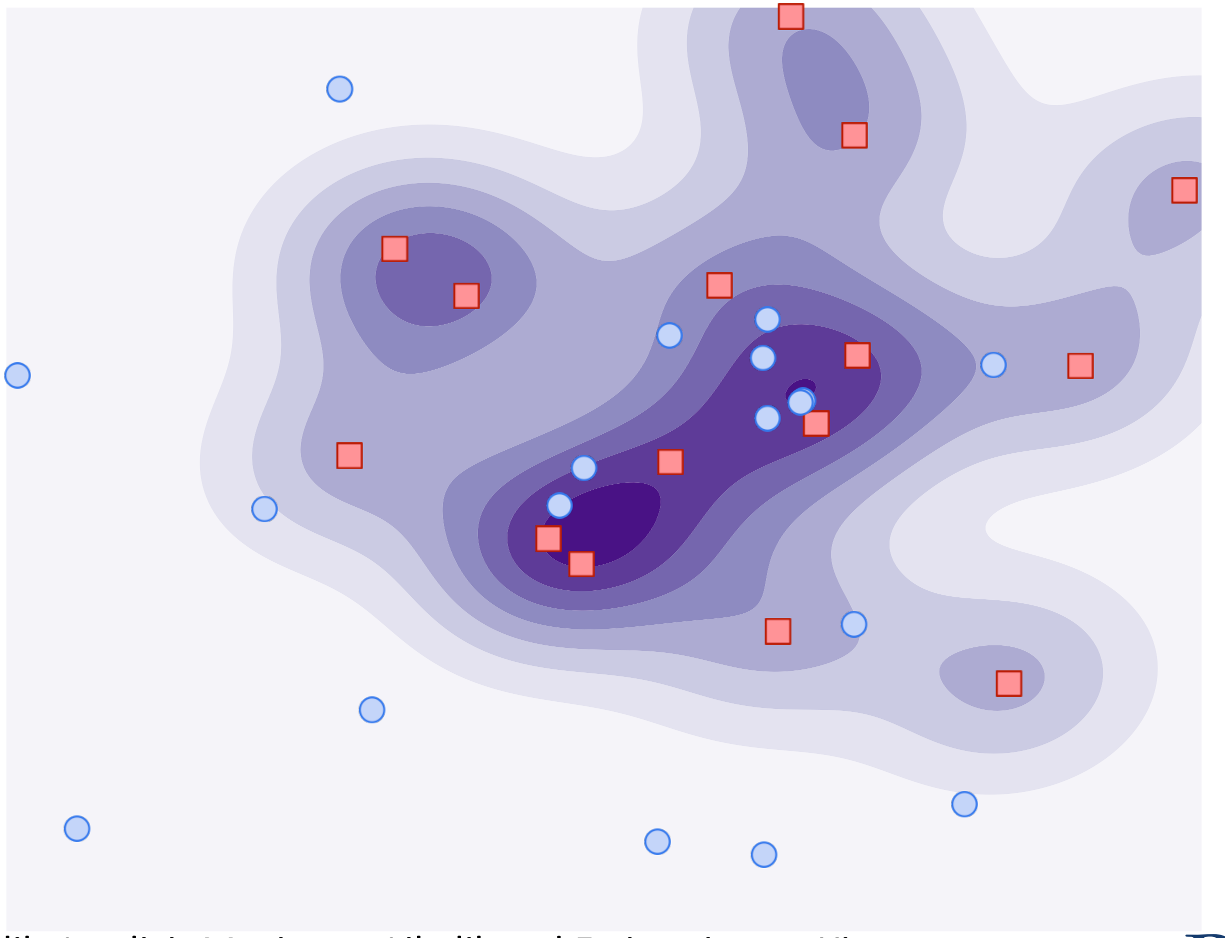
Move samples according to the gradient of the density estimate.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Maximize Likelihood of Samples Under Kernel Density Estimate

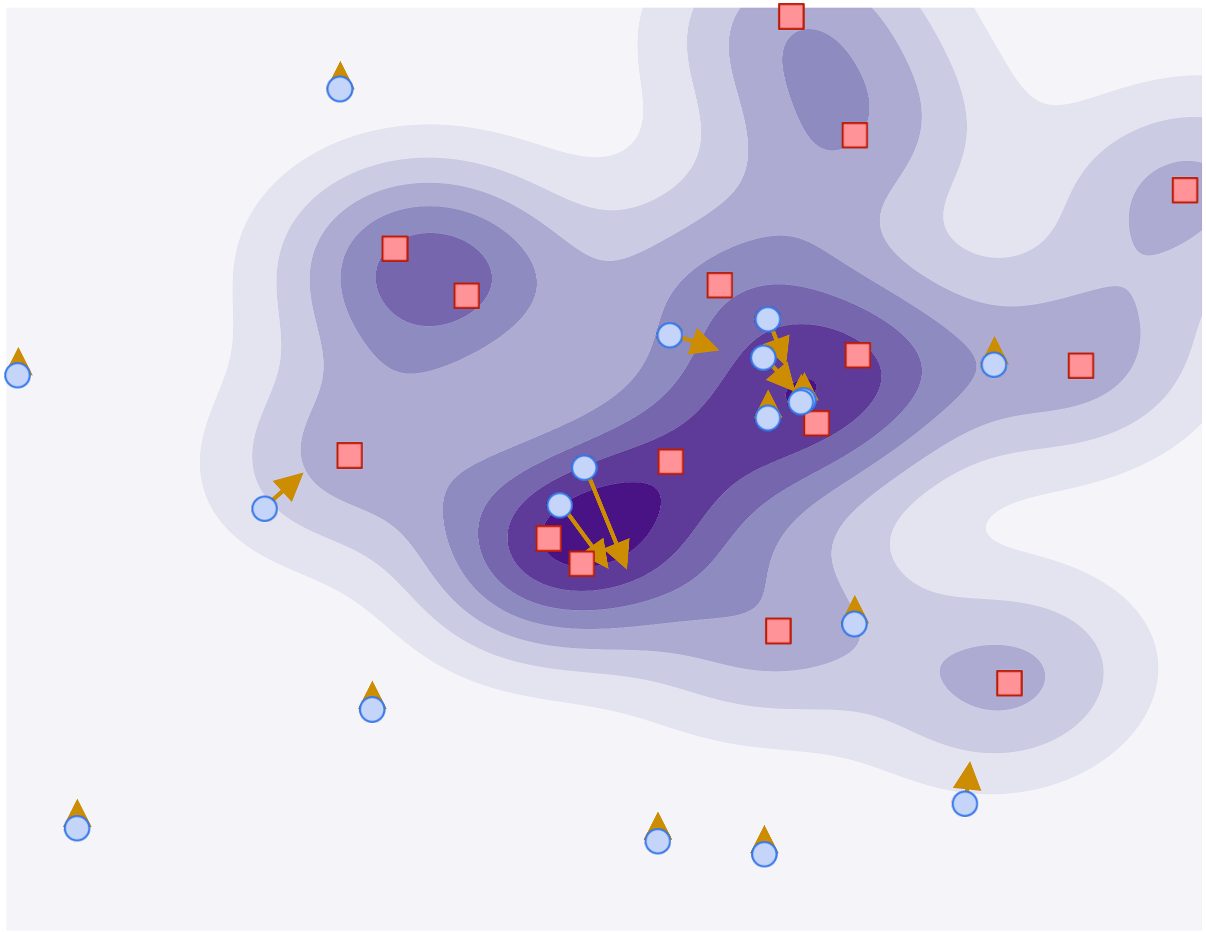


K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley UNIVERSITY OF CALIFORNIA

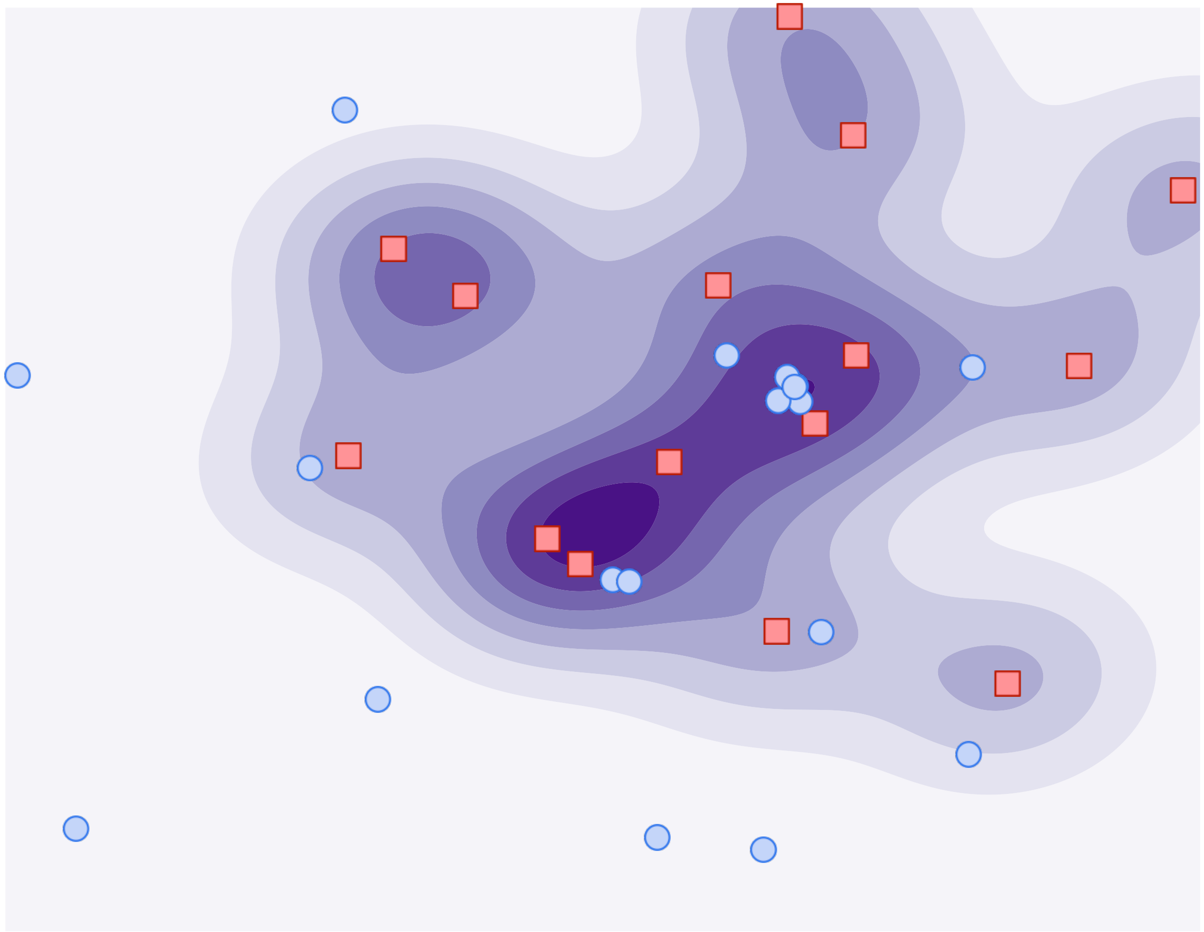
# Maximize Likelihood of Samples Under Kernel Density Estimate



K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

Berkeley UNIVERSITY OF CALIFORNIA

# Maximize Likelihood of Samples Under Kernel Density Estimate

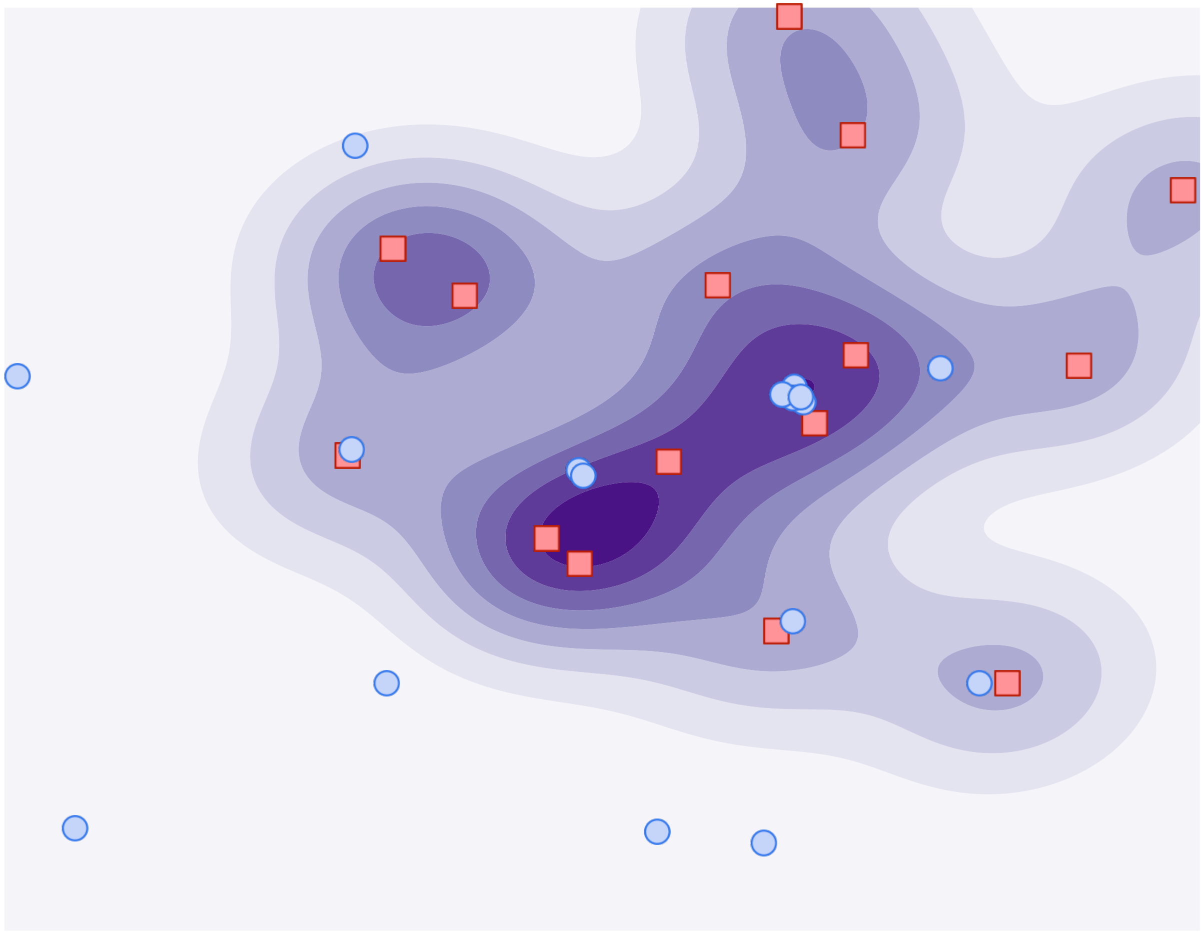


K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Maximize Likelihood of Samples Under Kernel Density Estimate

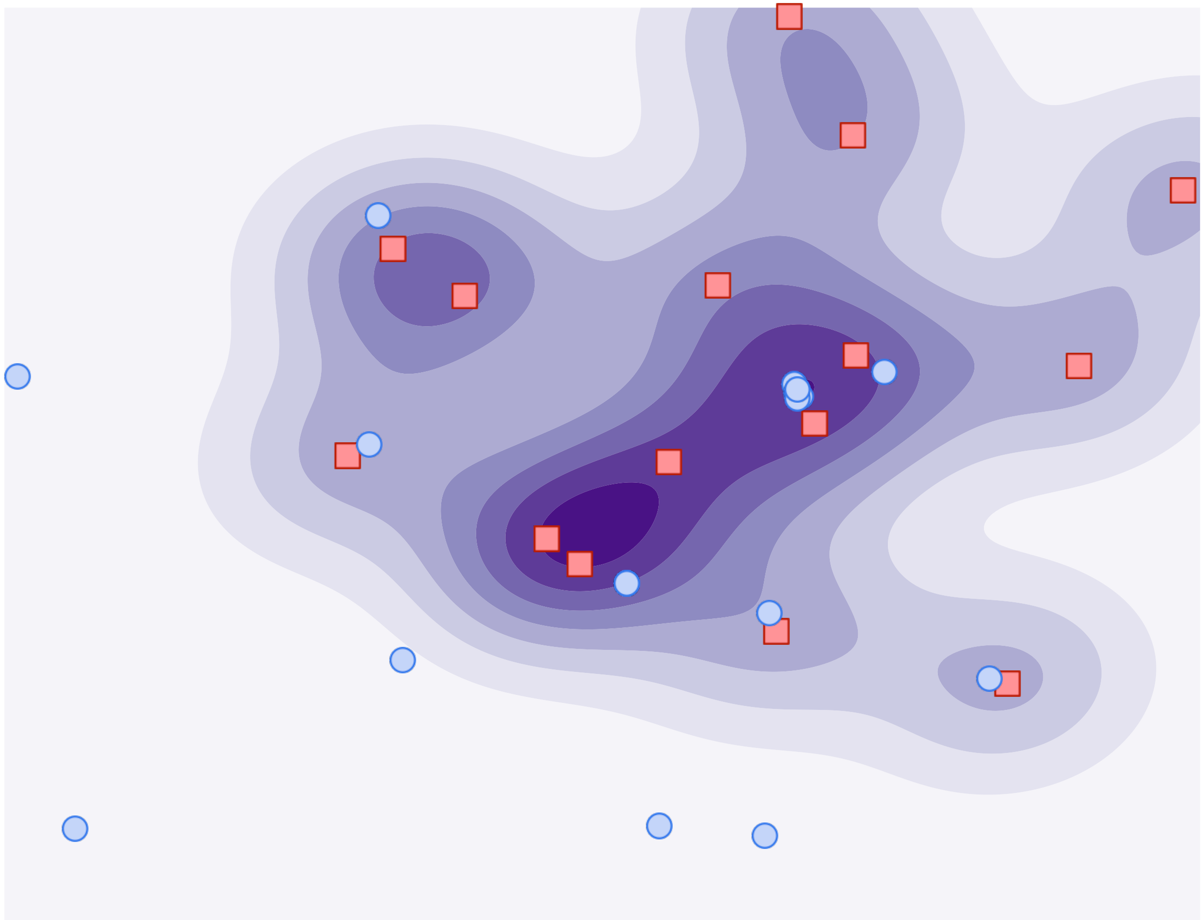


K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



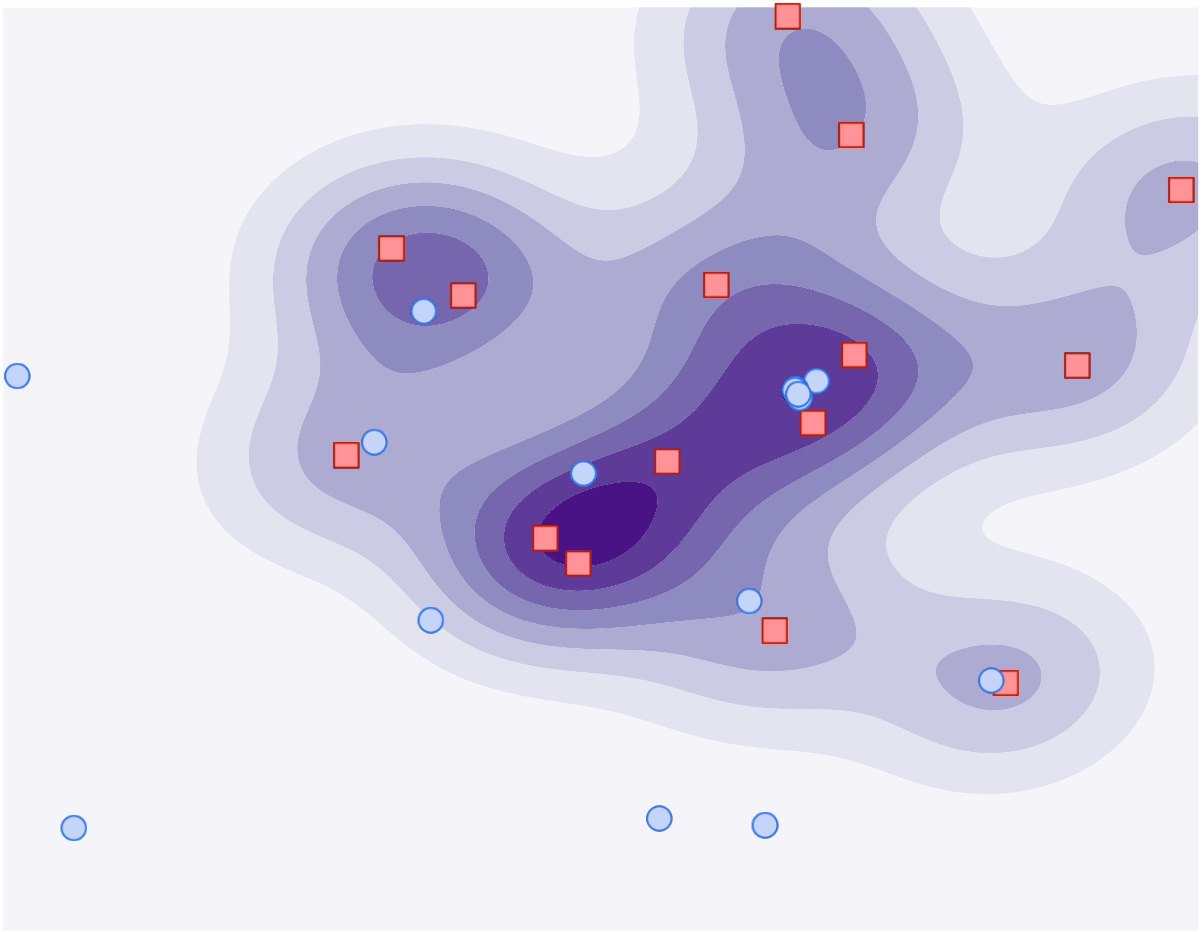
Berkeley
UNIVERSITY OF CALIFORNIA

# Maximize Likelihood of Samples Under Kernel Density Estimate



K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Maximize Likelihood of Samples Under Kernel Density Estimate

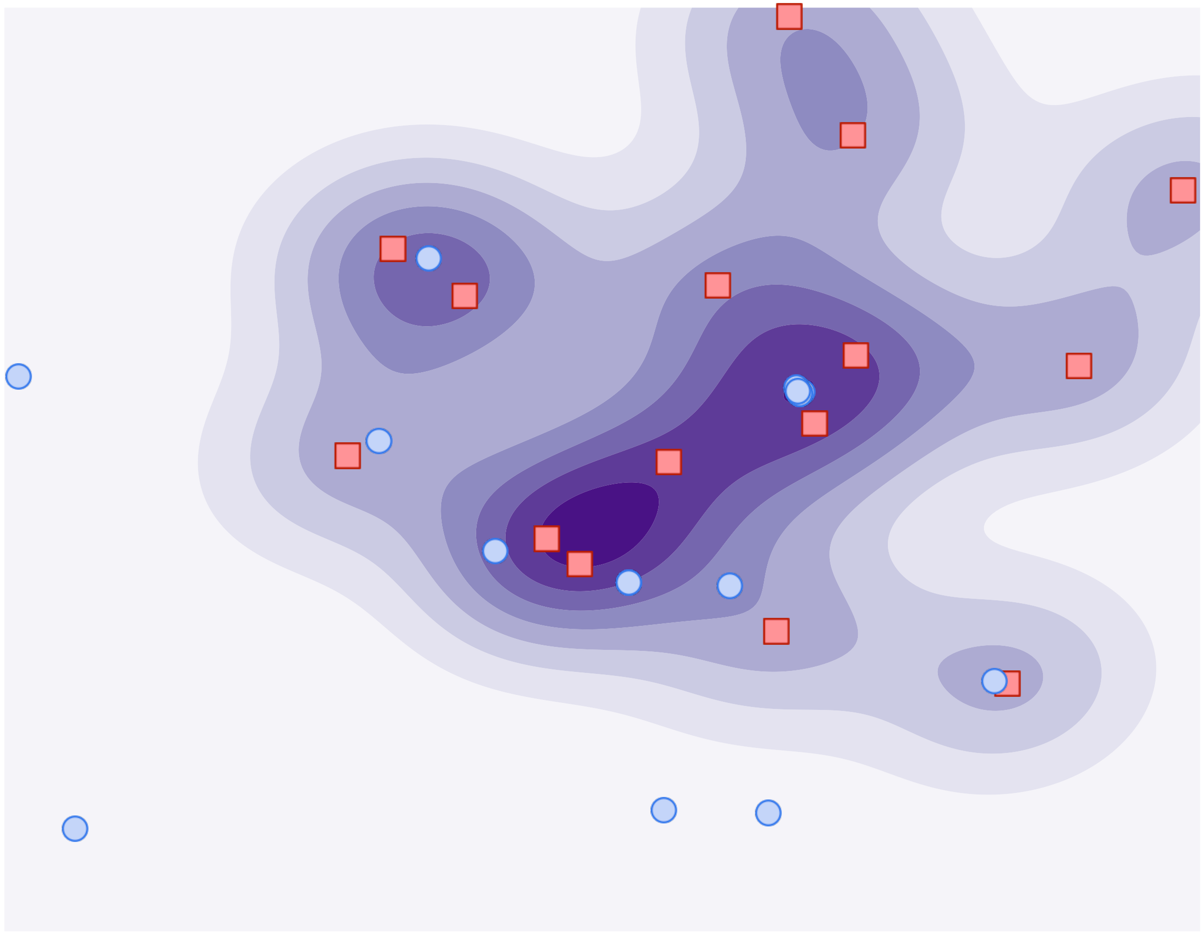


K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Maximize Likelihood of Samples Under Kernel Density Estimate

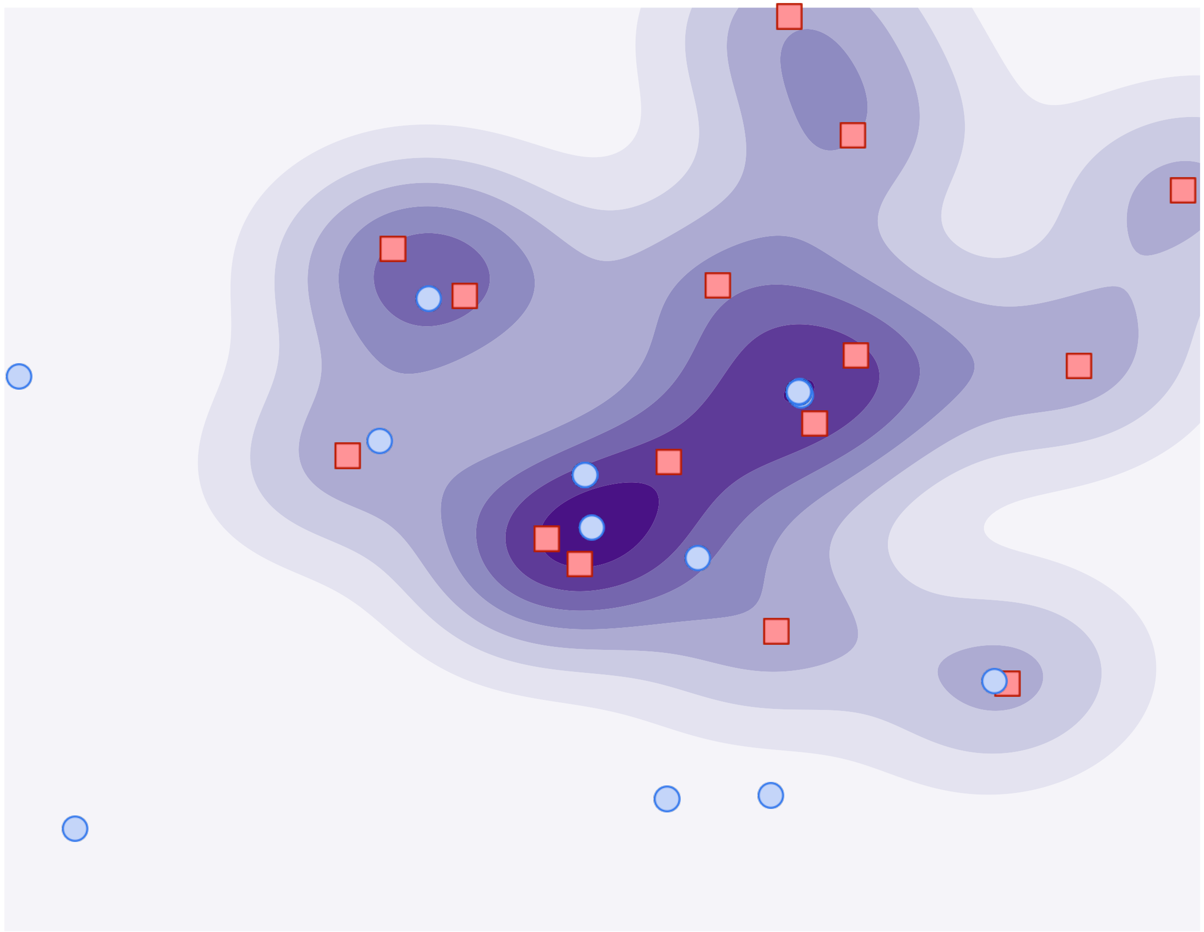


K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

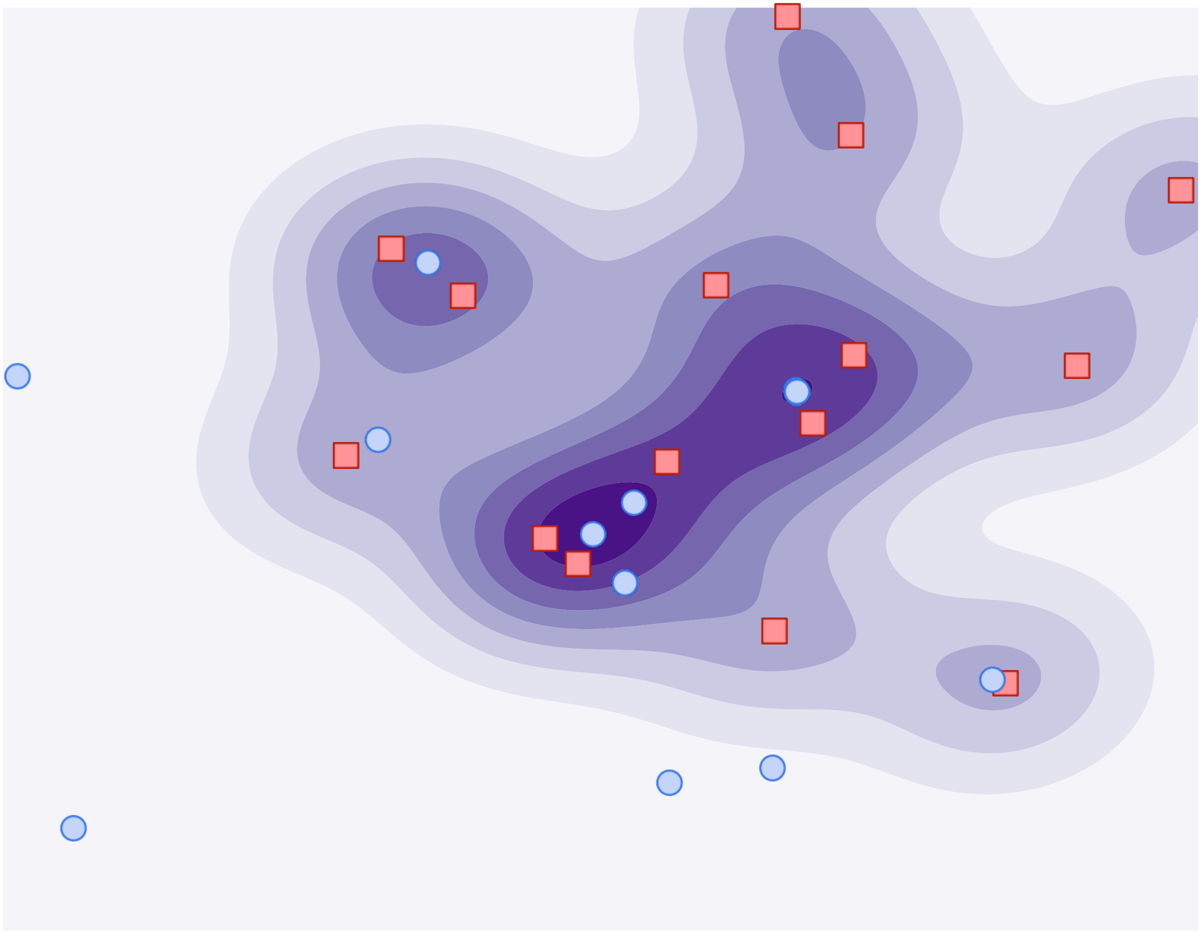
# Maximize Likelihood of Samples Under Kernel Density Estimate



K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

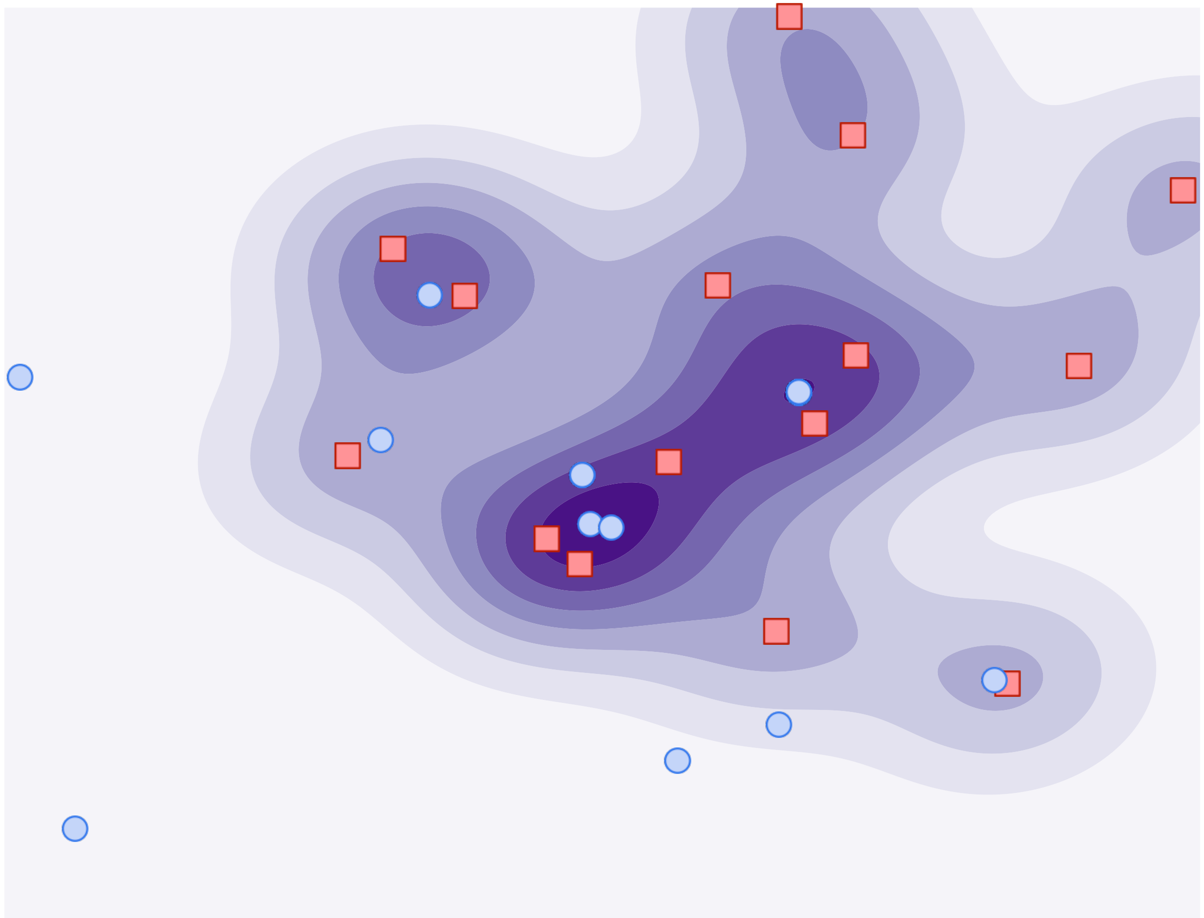
Berkeley
UNIVERSITY OF CALIFORNIA

# Maximize Likelihood of Samples Under Kernel Density Estimate



K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Maximize Likelihood of Samples Under Kernel Density Estimate

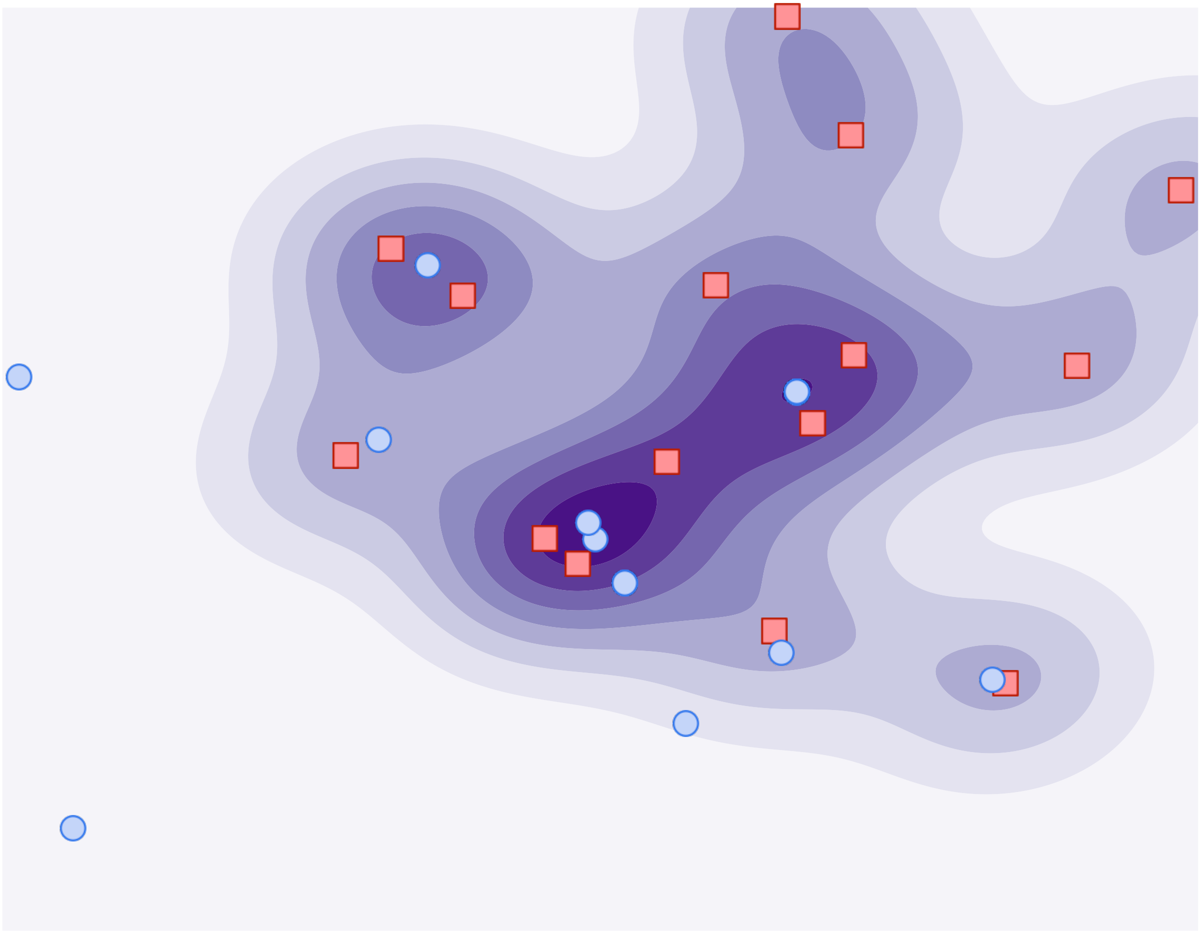


K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Maximize Likelihood of Samples Under Kernel Density Estimate

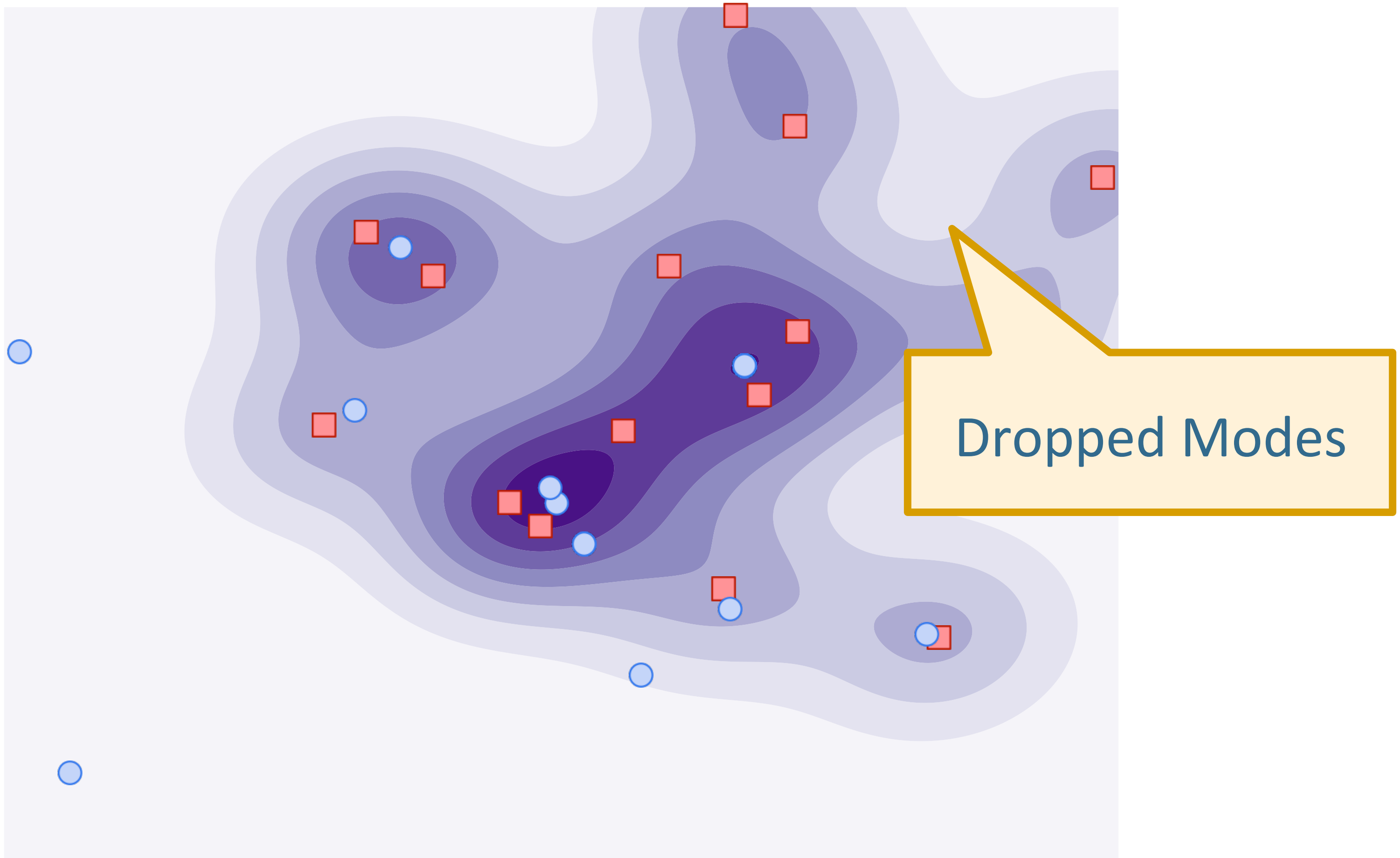


K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



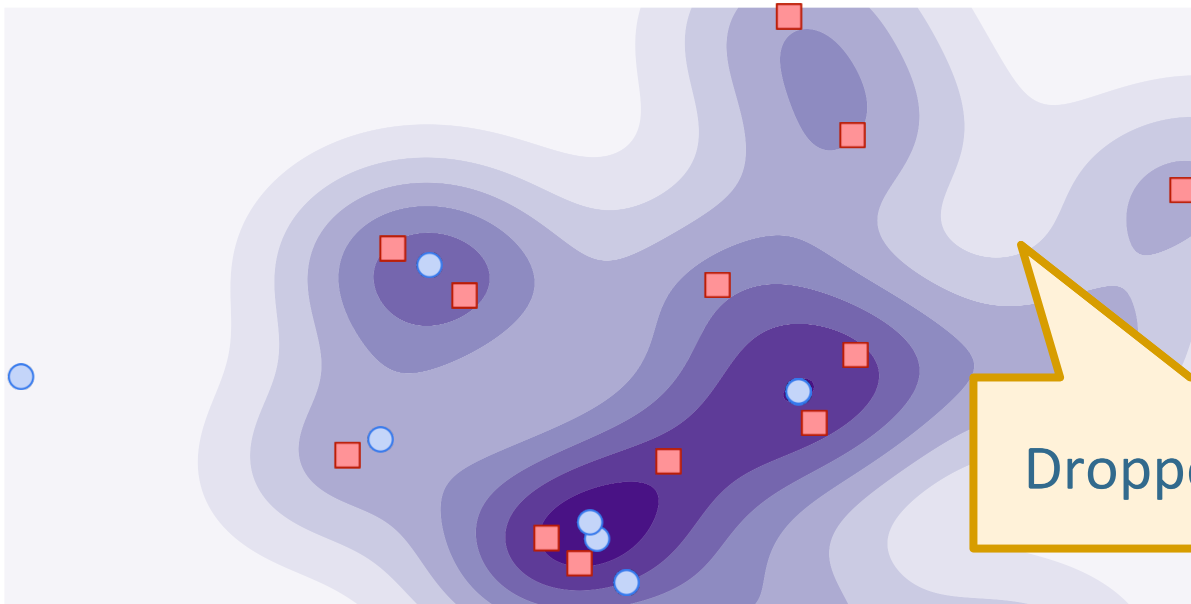
Berkeley
UNIVERSITY OF CALIFORNIA

# Maximize Likelihood of Samples Under Kernel Density Estimate




K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Maximize Likelihood of Samples Under Kernel Density Estimate



Dropped Modes

Again, each sample is essentially pushed towards the nearest mode. This is the opposite of IMLE.

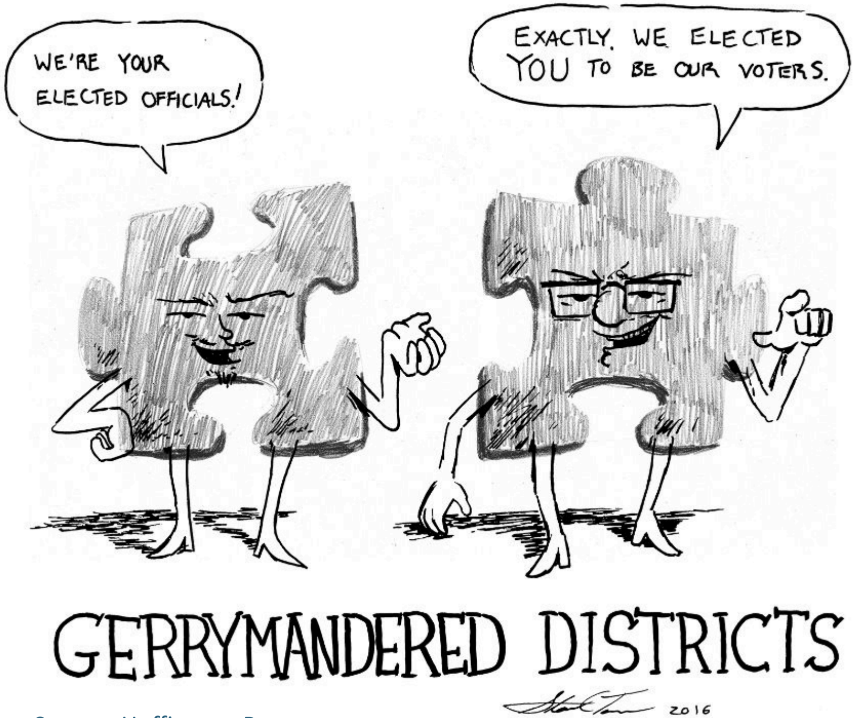
K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

Berkeley
UNIVERSITY OF CALIFORNIA

# Choosing Examples to Model

- In GANs, samples choose data examples; in IMLE, data examples choose samples – these are not the same! The former could drop modes, whereas the latter preserves all modes.
- This happens in real life too:

*"If an election is voters choosing politicians, gerrymandering is politicians choosing voters."*




Source: Huffington Post

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

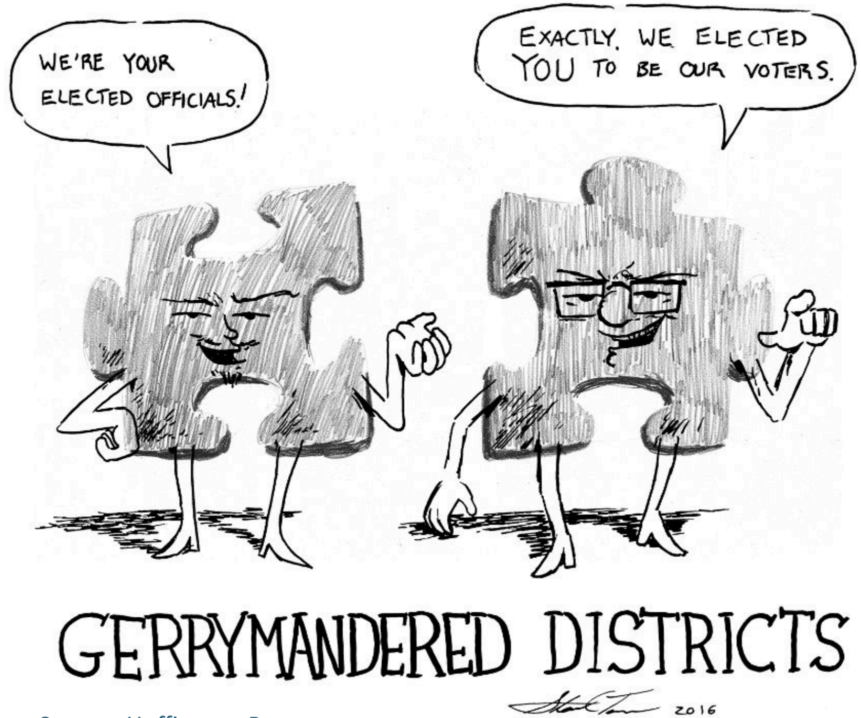
Berkeley UNIVERSITY OF CALIFORNIA

# Choosing Examples to Model

- In GANs, samples choose data examples; in IMLE, data examples choose samples – these are not the same! The former could drop modes, whereas the latter preserves all modes.
- This happens in real life too:

  *“If an election is voters choosing politicians, gerrymandering is politicians choosing voters.”*

- Who chooses whom matters!




Source: Huffington Post

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Problems of GANs

- In addition to mode dropping, GANs also suffer from other issues:
  - Vanishing gradients
  - Unstable training
- But according to the theory, there is a unique global pure-strategy Nash equilibrium, which corresponds to the model recovering the true data distribution.
- Why do the observed phenomena not match the theory?

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Problems of GANs

- Theoretical guarantee of GANs make three assumptions:
  - The discriminator has *infinite* modelling capacity
  - The number of samples from the data distribution must be *infinite*
  - Gradient descent-ascent converges to a global pure-strategy Nash equilibrium

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Problems of GANs

- Theoretical guarantee of GANs make three assumptions:
  - The discriminator has *infinite* modelling capacity
  - The number of samples from the data distribution must be *infinite*

Arora et al. (2017):

- Optimal generator could drop an exponential number of modes.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Problems of GANs

- Theoretical guarantee of GANs make three assumptions:
  - The discriminator has *infinite* modelling capacity
  - The number of samples from the data distribution must be *infinite*
  - Gradient descent-ascent converges to a global pure-strategy Nash equilibrium

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley UNIVERSITY OF CALIFORNIA

# Problems of GANs

- Theoretical guarantee of GANs make three assumptions:
  - The discriminator has *infinite* modelling capacity
  - The number of samples from the data distribution must be *infinite*

Arora et al. (2017):

- The JSD between the empirical data distribution and *any* continuous model distribution is always $\log 2$, which is the maximum possible value of JSD.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

Berkeley
UNIVERSITY OF CALIFORNIA

# Problems of GANs

- Theoretical guarantee of GANs make three assumptions:
  – The discriminator has *infinite* modelling capacity
  – The number of samples from the data distribution must be *infinite*

Sinn & Rawat (2017):

- Gradient w.r.t generator parameters given the optimal discriminator is zero almost everywhere.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Problems of GANs

- Theoretical guarantee of GANs make three assumptions:
  - The discriminator has *infinite* modelling capacity
  - The number of samples from the data distribution must be *infinite*

Arora et al. (2017):

- Minimizing JSD between the empirical data distribution and the model does not necessarily recover the true data distribution.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

Berkeley
UNIVERSITY OF CALIFORNIA

# Problems of GANs

- Theoretical guarantee of GANs make three assumptions:
  - The discriminator has *infinite* modelling capacity
  - The number of samples from the data distribution must be *infinite*
  - Gradient descent-ascent converges to a global pure-strategy Nash equilibrium

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Problems of GANs

Arora et al. (2017):

- In the parametric setting, a global pure-strategy Nash equilibrium may not exist.

Gilboa & Zemel (1989):

- Finding it would be NP-hard.

– ~~Gradient descent-ascent converges to a global pure-strategy Nash equilibrium~~

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Problems of GANs

Mescheder et al. (2017):

- Gradient descent-ascent will not necessarily find a local pure-strategy Nash equilibrium and is only guaranteed to find a stationary point (e.g. when the discriminator is too powerful and winning the game against the generator).
- Convergence could be slow.

– Gradient descent-ascent converges to a global pure-strategy Nash equilibrium

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# IMLE Overcomes These Problems

**No More Mode Collapse**

- Each data example has a nearby sample.

**No More Vanishing Gradients**

- Gradient only becomes zero when distance to nearest sample is zero.

**No More Training Instability**

- Simple minimization problem – no adversarial formulation.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Agenda

1. Implicit Maximum Likelihood Estimation
2. Comparison to Generative Adversarial Nets (GANs)
3. *Why Maximum Likelihood*
4. Equivalence to Maximum Likelihood
5. Fast Nearest Neighbour Search
6. Applications to Conditional Image Synthesis

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Why Maximum Likelihood

# Reverse KL vs. Standard KL

- Huszar (2015): "How (Not) to Train Your Generative Model"

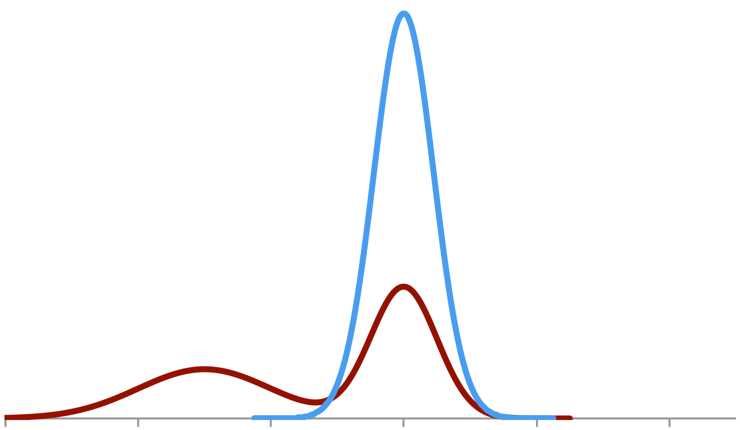
right training objective when the goal is to generate realistic samples. Maximum likelihood can be thought of as minimising the Kullback-Leibler divergence $KL[P\|Q]$ between the real data distribution $P$ and the probabilistic model $Q$. We present a model that suggests generative models should instead be trained to minimise $KL[Q\|P]$, the Kullback-Leibler divergence in the opposite direction. The differences between minimising $KL[P\|Q]$ and $KL[Q\|P]$ are well understood, and explain the observed undesirable behaviour in autoregressive sequence models.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

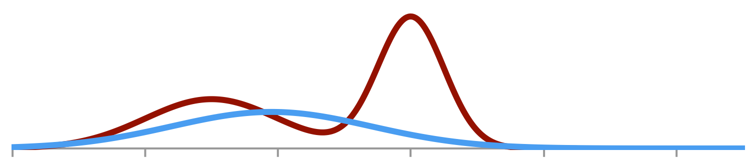


Berkeley UNIVERSITY OF CALIFORNIA

# Reverse KL vs. Standard KL



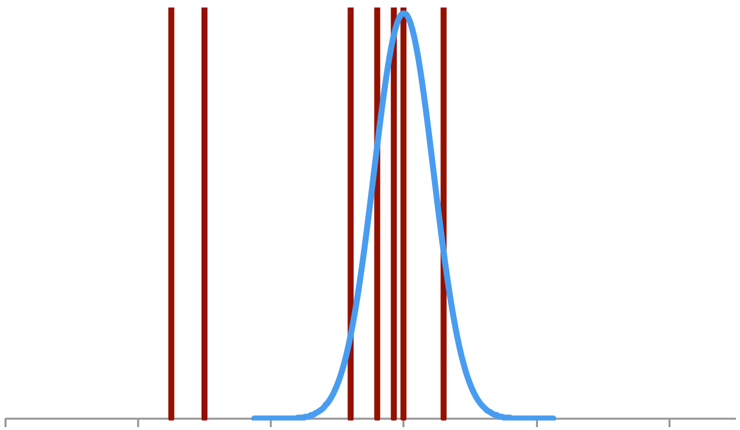
$$D_{KL}\left(p_\theta \| p_{\text{data}}\right) = \int_{-\infty}^{\infty} p_\theta(x) \log \frac{p_\theta(x)}{p_{\text{data}}(x)} dx$$

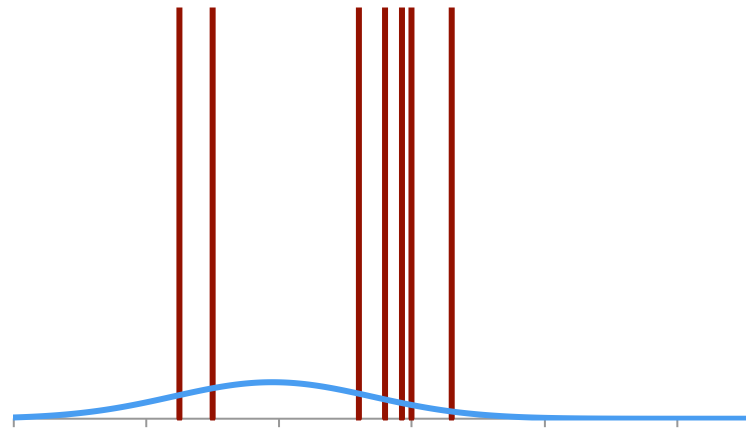


$$D_{KL}\left(p_{\text{data}} \| p_\theta\right) = \int_{-\infty}^{\infty} p_{\text{data}}(x) \log \frac{p_{\text{data}}(x)}{p_\theta(x)} dx$$

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# But we only have the *empirical* data distribution…





$$D_{KL}\left(p_\theta \,\|\, p_{\text{data}}\right) = \int_{-\infty}^{\infty} p_\theta(x) \log \frac{p_\theta(x)}{p_{\text{data}}(x)} dx$$

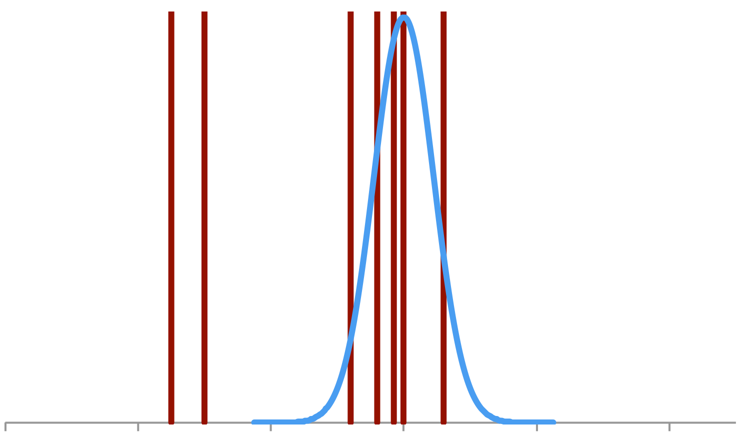
$$D_{KL}\left(p_{\text{data}} \,\|\, p_\theta\right) = \int_{-\infty}^{\infty} p_{\text{data}}(x) \log \frac{p_{\text{data}}(x)}{p_\theta(x)} dx$$

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Is Reverse KL even defined in this setting?

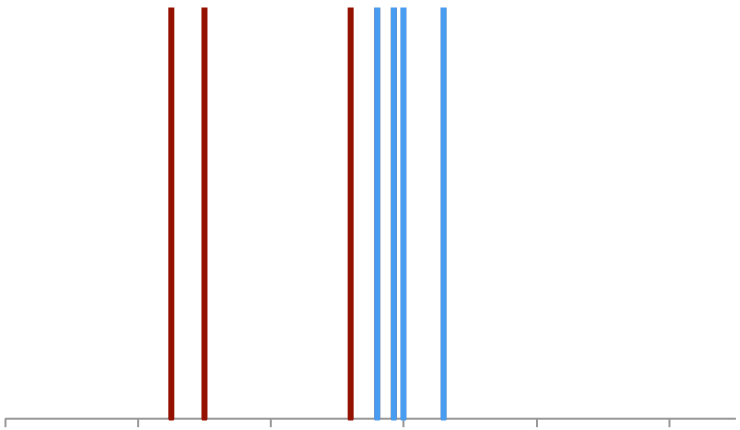


$$D_{KL}\left(p_\theta \| p_{\text{data}}\right) = \int_{-\infty}^{\infty} p_\theta(x) \log \frac{p_\theta(x)}{p_{\text{data}}(x)} dx$$

- Recall:
  - For $D_{KL}(p\|q)$ to be finite, $p$ must be absolutely continuous w.r.t. $q$, i.e.:
  - If $q(x) = 0$, $p(x) = 0$
- For reverse KL to be finite, support of the model distribution must be a subset of the support of the empirical data distribution.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Is Reverse KL even defined in this setting?



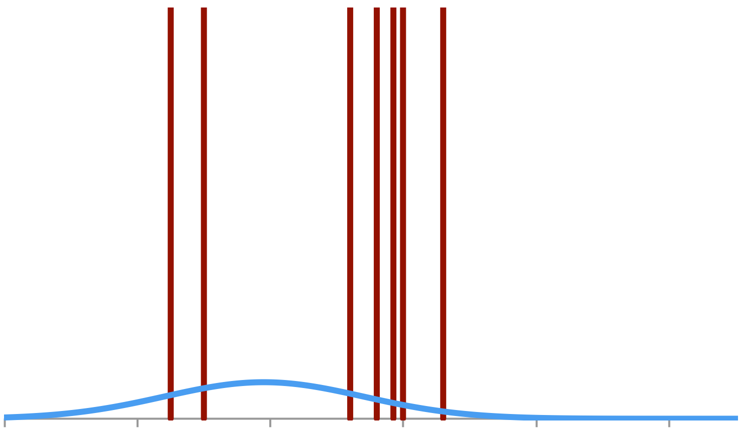
$$D_{KL}\left(p_\theta \| p_{\text{data}}\right) = \int_{-\infty}^{\infty} p_\theta(x) \log \frac{p_\theta(x)}{p_{\text{data}}(x)} dx$$

- Recall:
  - For $D_{KL}(p\|q)$ to be finite, $p$ must be absolutely continuous w.r.t. $q$, i.e.:
  - If $q(x) = 0$, $p(x) = 0$
- For reverse KL to be finite, support of the model distribution must be a subset of the support of the empirical data distribution.
- Effectively, generalization is prohibited.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Standard KL

- For standard KL divergence to be finite, support of the empirical data distribution must be a subset of the support of the model distribution.
- Model must capture all data examples – mode dropping is impossible.



$$D_{KL}\left(p_{\text{data}} \| p_{\theta}\right) = \int_{-\infty}^{\infty} p_{\text{data}}(x) \log \frac{p_{\text{data}}(x)}{p_{\theta}(x)} dx$$

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley UNIVERSITY OF CALIFORNIA

# Reverse KL doesn't make sense

- But what if the model cannot capture all data examples?
- Two options:
    1. Let the model pick and choose which data examples it wants to model and model only those examples (drop modes)
    2. Come up with more expressive models that can model all data examples (advance science)
- Why is Option 2 better?

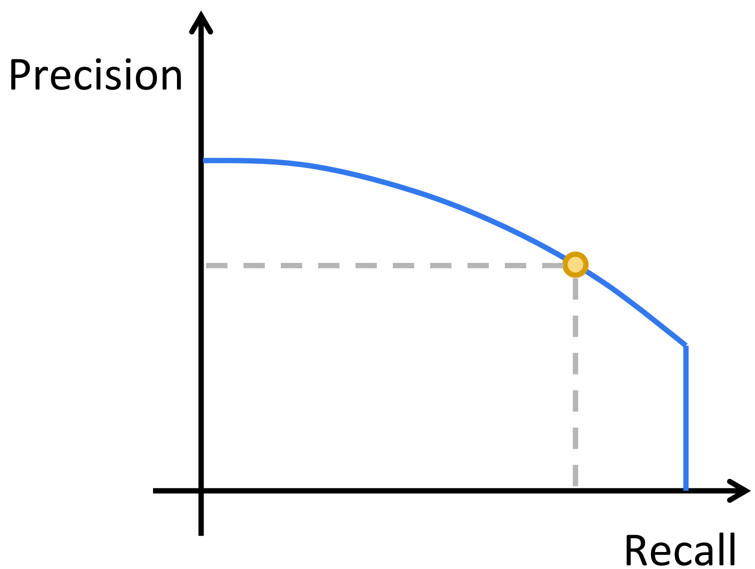
K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Precision and Recall

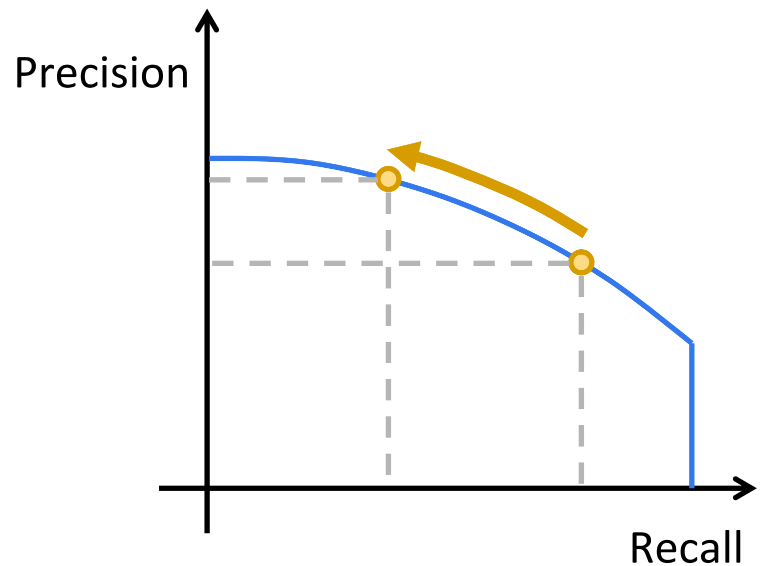
- Precision: How accurately we model the modes that are modelled (reflected by sample quality)
- Recall: How many modes we model (reflected by sample diversity)
- If we keep the model fixed, then there is a tradeoff between precision and recall.




K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Precision and Recall

- An improvement in precision can be due to:
  1. A movement along the curve (trade off recall for precision)



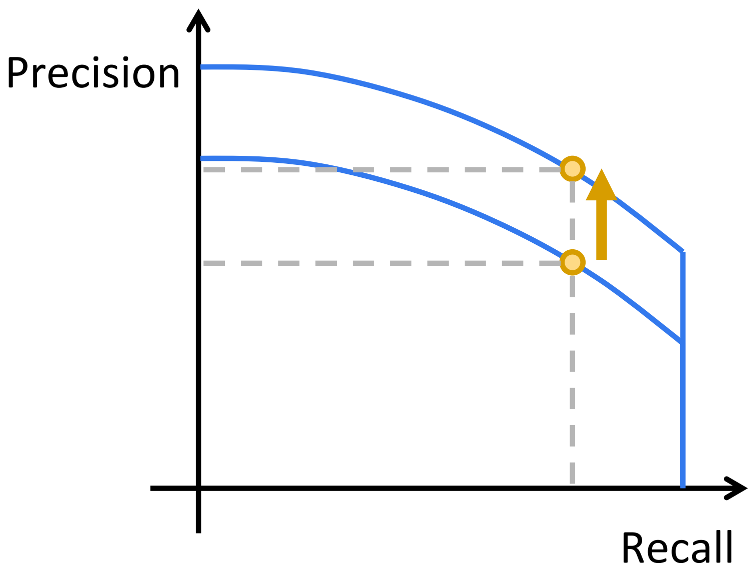

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Precision and Recall

- An improvement in precision can be due to:
  1. A movement along the curve (trade off recall for precision)
  2. A shift in the curve (improvement at all levels of recall)



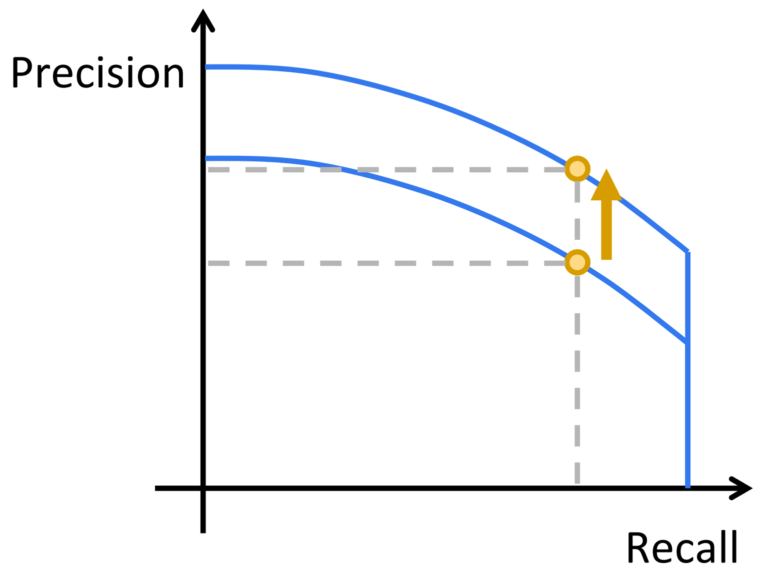

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Precision and Recall

- An improvement in precision can be due to:
  1. A movement along the curve (trade off recall for precision)
  2. A shift in the curve (improvement at all levels of recall)
- Either/both may take place; unfortunately we don't know which actually took place because we have no control over the modes that are dropped.




K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley UNIVERSITY OF CALIFORNIA

# Implication on the Trajectory of Research

- Sample diversity is not easy to measure, but sample quality is. Consequently, most work in recent years focused on sample quality.
- This is inherently biased towards precision, at the expense of recall.
- This is not good, because we don't how much sample diversity we are losing in the pursuit of sample quality.

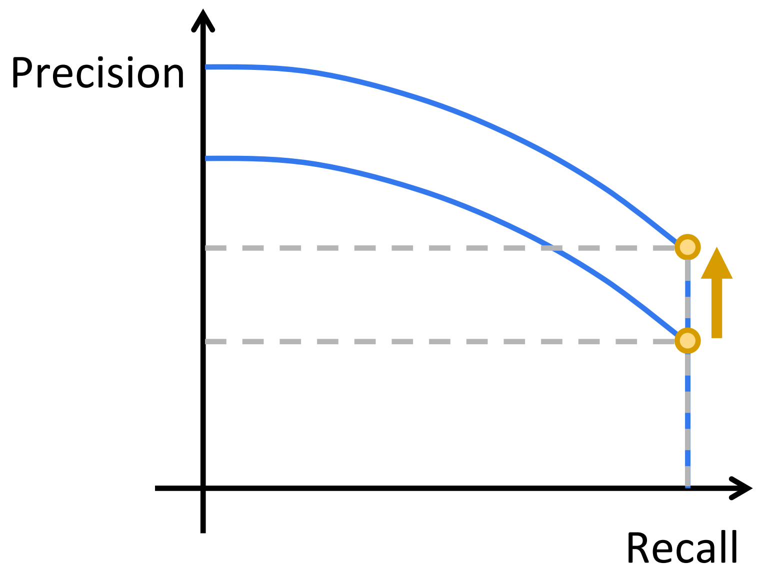
K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# A Better Way

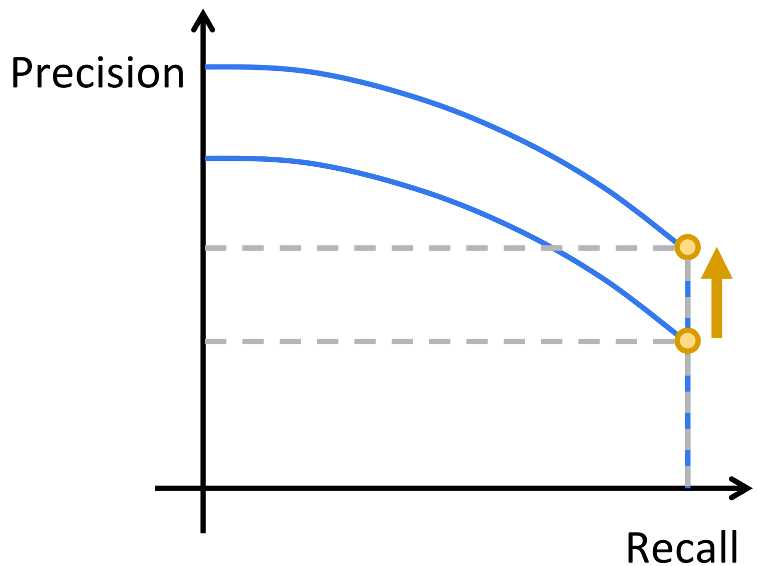
- What if we insist on maximizing likelihood?
  - Likelihood is a product of the densities at all data examples, so if one data example is assigned zero density, likelihood would be zero.
  - So, maximum likelihood prevents mode dropping.
  - Hence, full recall is guaranteed.




K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# A Better Way

- When full recall is guaranteed, improving precision/sample quality is guaranteed to improve the model.
- This is why sample quality used to be correlated with log-likelihoods in pre-2014 papers and why sample quality has traditionally been used to measure the performance of generative models.




K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Agenda

1. Implicit Maximum Likelihood Estimation
2. Comparison to Generative Adversarial Nets (GANs)
3. Why Maximum Likelihood
4. *Equivalence to Maximum Likelihood*
5. Fast Nearest Neighbour Search
6. Applications to Conditional Image Synthesis

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Equivalence to Maximum Likelihood

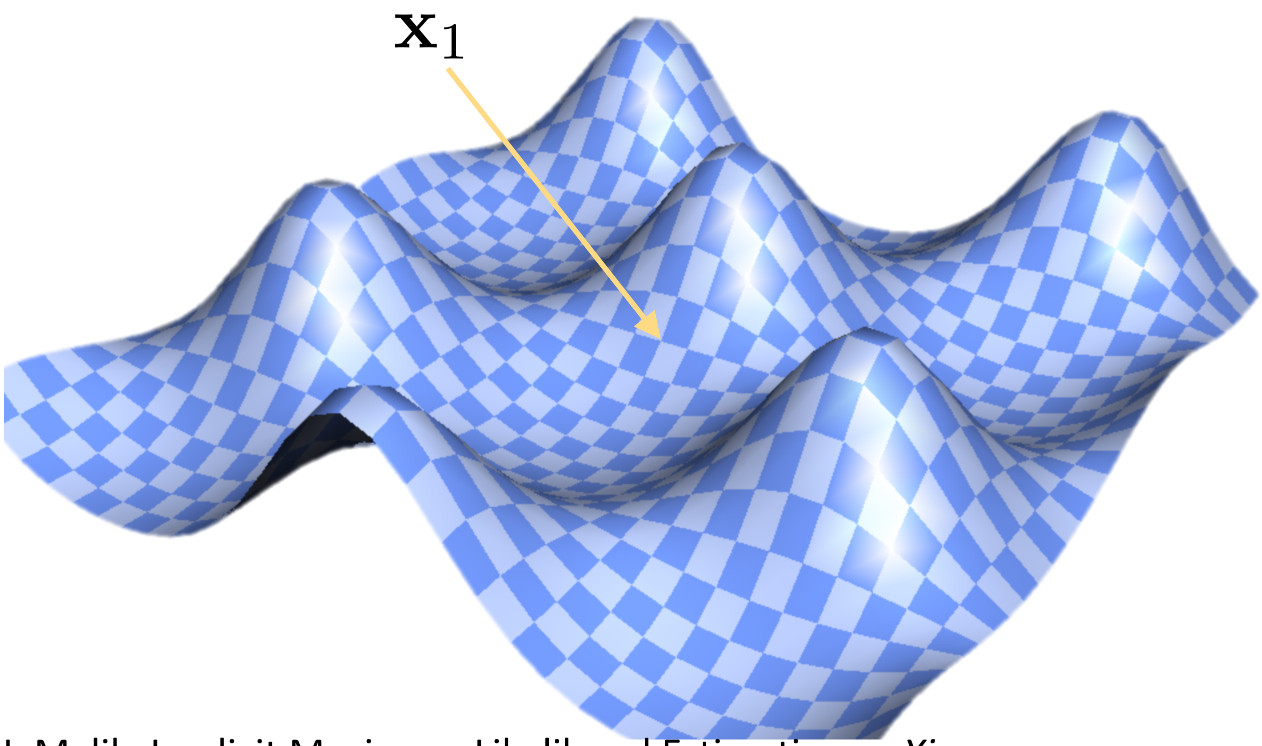
# Equivalence to Maximum Likelihood

- For simplicity, suppose we only have one data example $\mathbf{x}_1$ and only generate one sample $\tilde{\mathbf{x}}_1^{\theta}$ .

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Equivalence to Maximum Likelihood

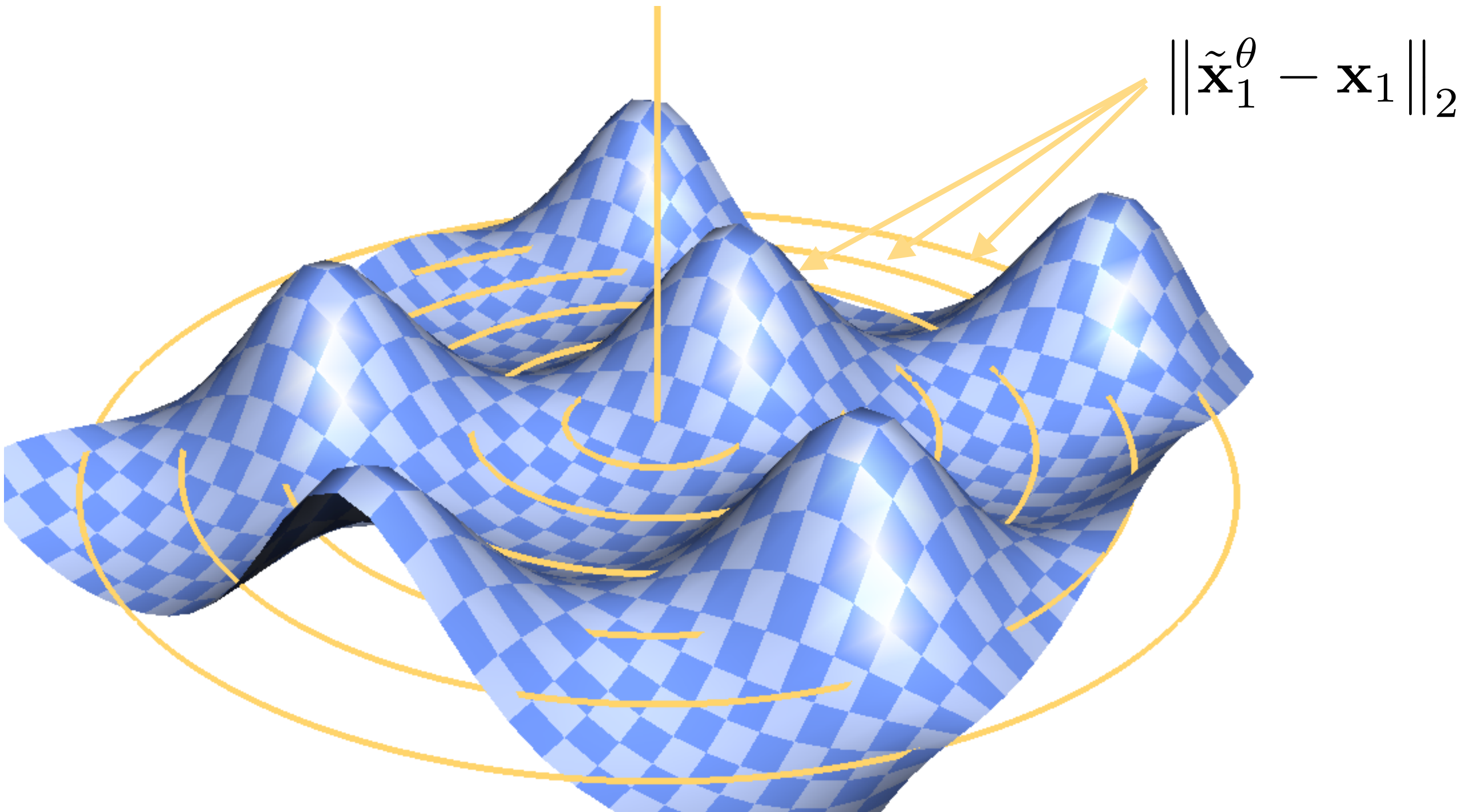
- Model density:




K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Equivalence to Maximum Likelihood

- How is the random variable $\left\|\tilde{\mathbf{x}}_1^\theta - \mathbf{x}_1\right\|_2$ distributed?

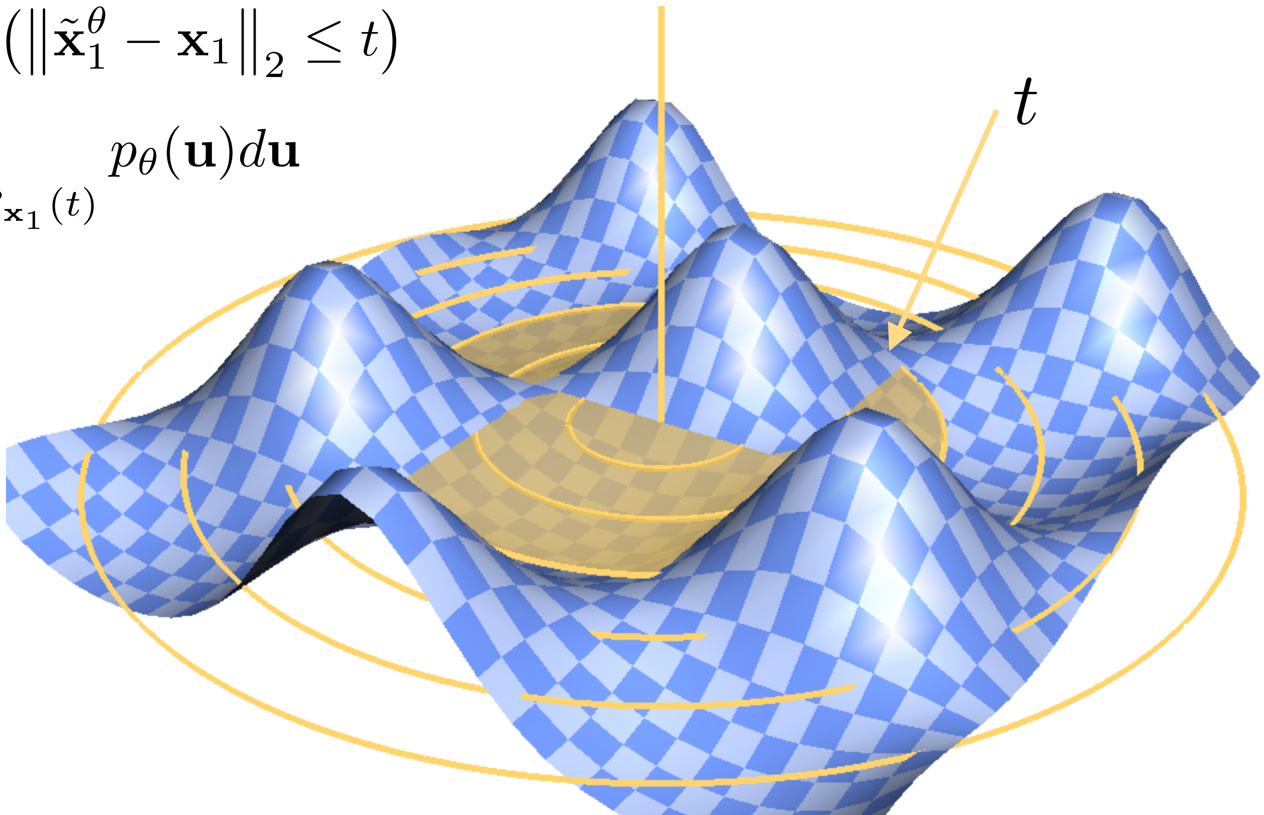



K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Equivalence to Maximum Likelihood

- How is the random variable $\left\|\tilde{\mathbf{x}}_1^\theta - \mathbf{x}_1\right\|_2$ distributed?

$$\Pr\left(\left\|\tilde{\mathbf{x}}_1^\theta - \mathbf{x}_1\right\|_2 \leq t\right)$$
$$= \int_{B_{\mathbf{x}_1}(t)} p_\theta(\mathbf{u}) d\mathbf{u}$$

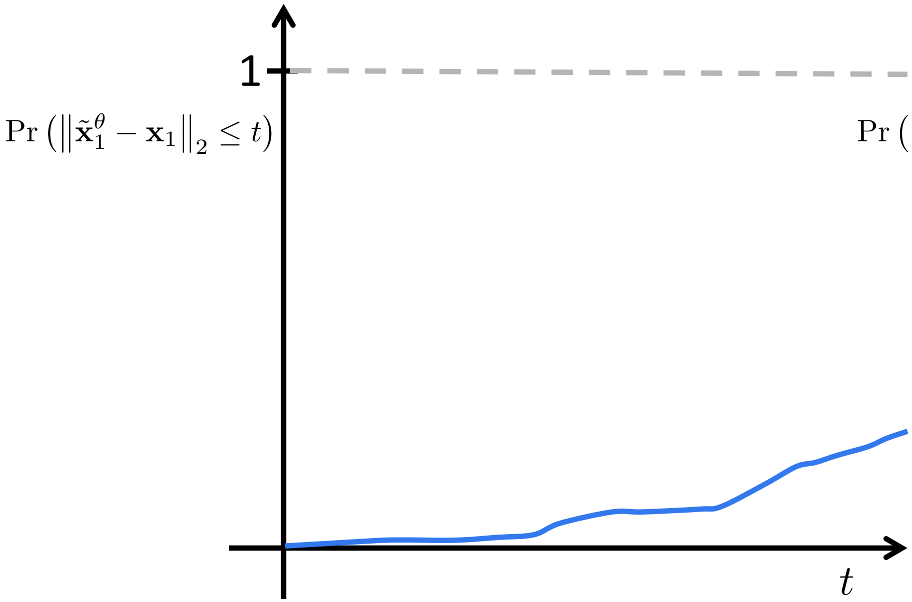



K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

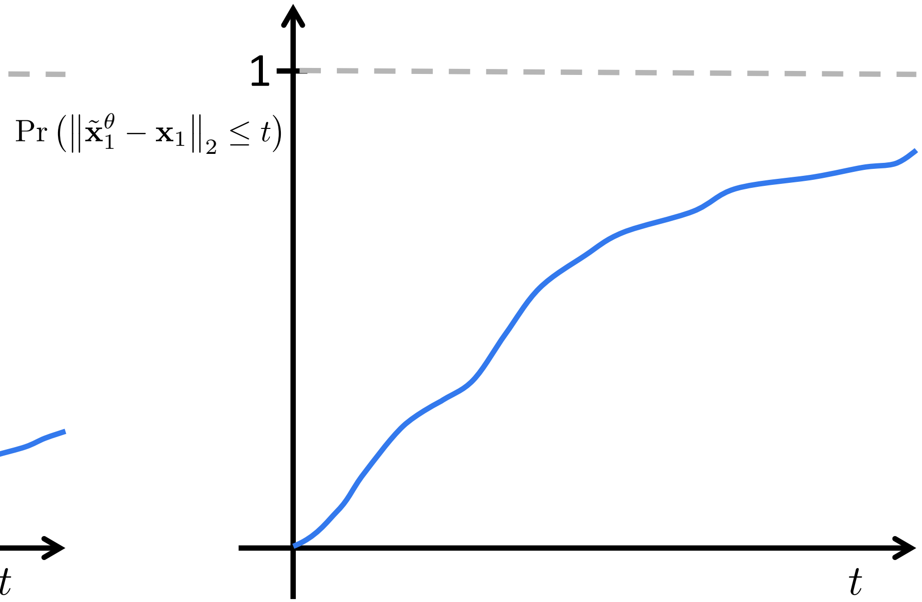
Berkeley
UNIVERSITY OF CALIFORNIA

# Equivalence to Maximum Likelihood

- The CDF of $\left\|\tilde{\mathbf{x}}_1^\theta - \mathbf{x}_1\right\|_2$ looks like:




When likelihood is low



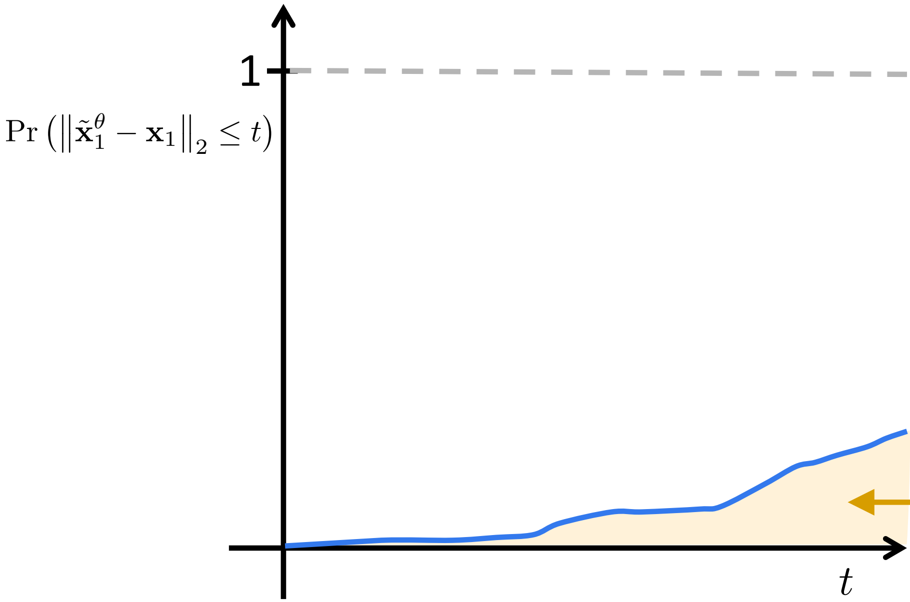

When likelihood is high

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

Berkeley UNIVERSITY OF CALIFORNIA

# Equivalence to Maximum Likelihood

- Likelihood is maximized when the CDF grows quickly.

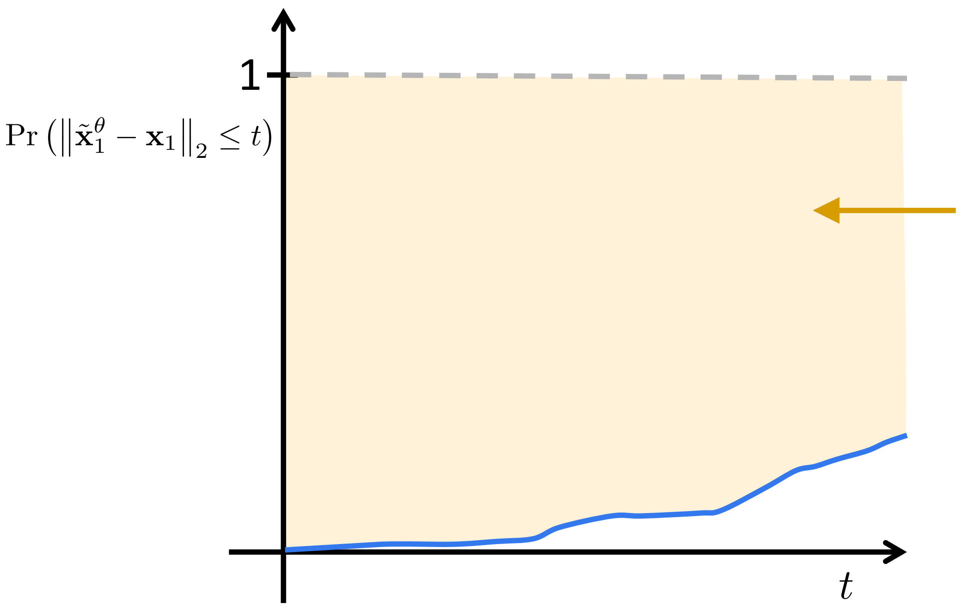



To make the CDF grow quickly, we can maximize the area under the curve.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Equivalence to Maximum Likelihood

- Likelihood is maximized when the CDF grows quickly.




To make the CDF grow quickly, we can minimize the area above the curve, which is the the same as the minimizing the area under the curve of the function

$$1 - \Pr\left(\left\|\tilde{\mathbf{x}}_1^\theta - \mathbf{x}_1\right\|_2 \le t\right)$$
$$= \Pr\left(\left\|\tilde{\mathbf{x}}_1^\theta - \mathbf{x}_1\right\|_2 > t\right).$$

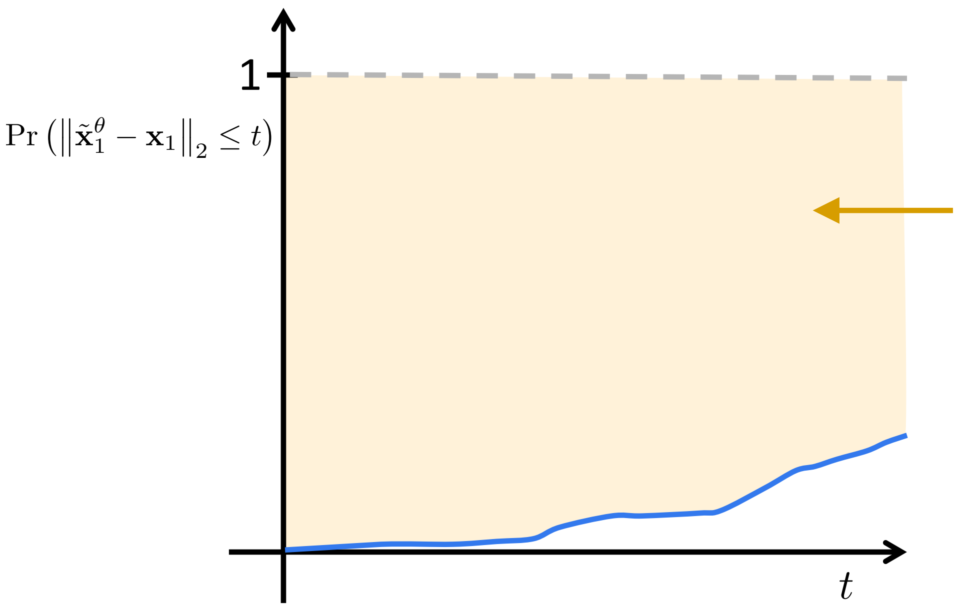
K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley UNIVERSITY OF CALIFORNIA

# Equivalence to Maximum Likelihood

- Likelihood is maximized when the CDF grows quickly.




The area of the shaded region is:

$$\int_0^\infty \Pr\left(\left\|\tilde{\mathbf{x}}_1^\theta - \mathbf{x}_1\right\|_2 > t\right) dt$$

Recall that for any non-negative random variable $X$,

$$\mathbb{E}\left[X\right] = \int_0^\infty \Pr\left(X > t\right) dt$$

So, the area is exactly $\mathbb{E}\left[\left\|\tilde{\mathbf{x}}_1^\theta - \mathbf{x}_1\right\|_2\right]$.

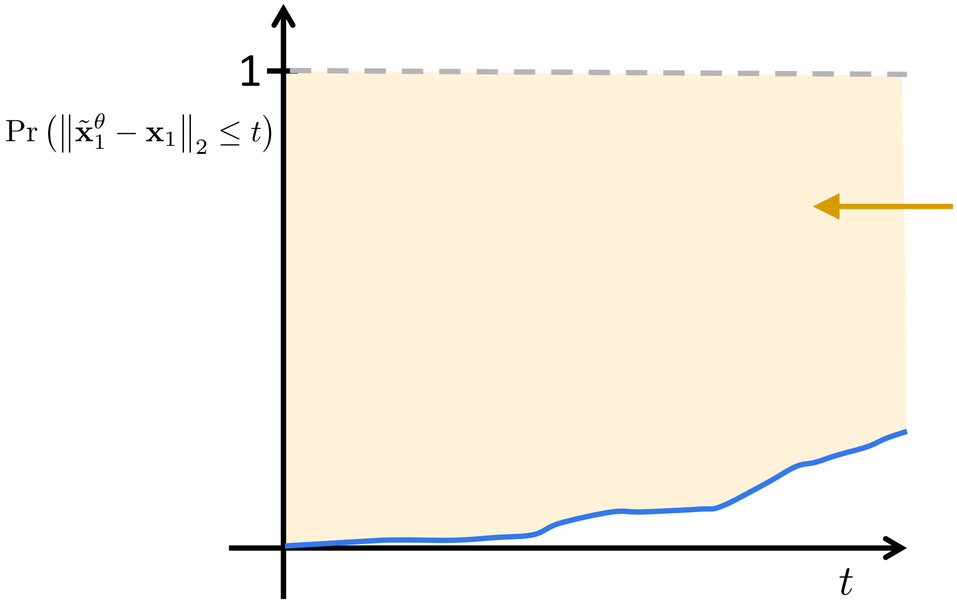
K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley UNIVERSITY OF CALIFORNIA

# Equivalence to Maximum Likelihood

- Likelihood is maximized when the CDF grows quickly.




Hence, minimizing $\mathbb{E}\left[\left\|\tilde{\mathbf{x}}_1^\theta - \mathbf{x}_1\right\|_2\right]$
is equivalent to maximizing likelihood.

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley UNIVERSITY OF CALIFORNIA

# Main Result

- Let $\tilde{\mathbf{x}}_1^\theta, \ldots, \tilde{\mathbf{x}}_m^\theta \sim P_\theta$ be i.i.d. samples and
$R_i^\theta := \min_{j \in [m]} \left\| \tilde{\mathbf{x}}_j^\theta - \mathbf{x}_i \right\|_2^2$
- If the maximum likelihood solution $\theta^*$ is unique, then under some conditions,

$$\arg\max_\theta \sum_{i=1}^n \log p_\theta(\mathbf{x}_i) = \arg\min_\theta \sum_{i=1}^n \frac{\mathbb{E}\left[R_i^\theta\right]}{\Psi'(p_{\theta^*}(\mathbf{x}_i))p_{\theta^*}(\mathbf{x}_i)}$$

- Furthermore, if $p_{\theta^*}(\mathbf{x}_1) = \cdots = p_{\theta^*}(\mathbf{x}_n)$,

$$\arg\max_\theta \sum_{i=1}^n \log p_\theta(\mathbf{x}_i) = \arg\min_\theta \sum_{i=1}^n \mathbb{E}\left[R_i^\theta\right]$$

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Results

# Estimated Log-Likelihood

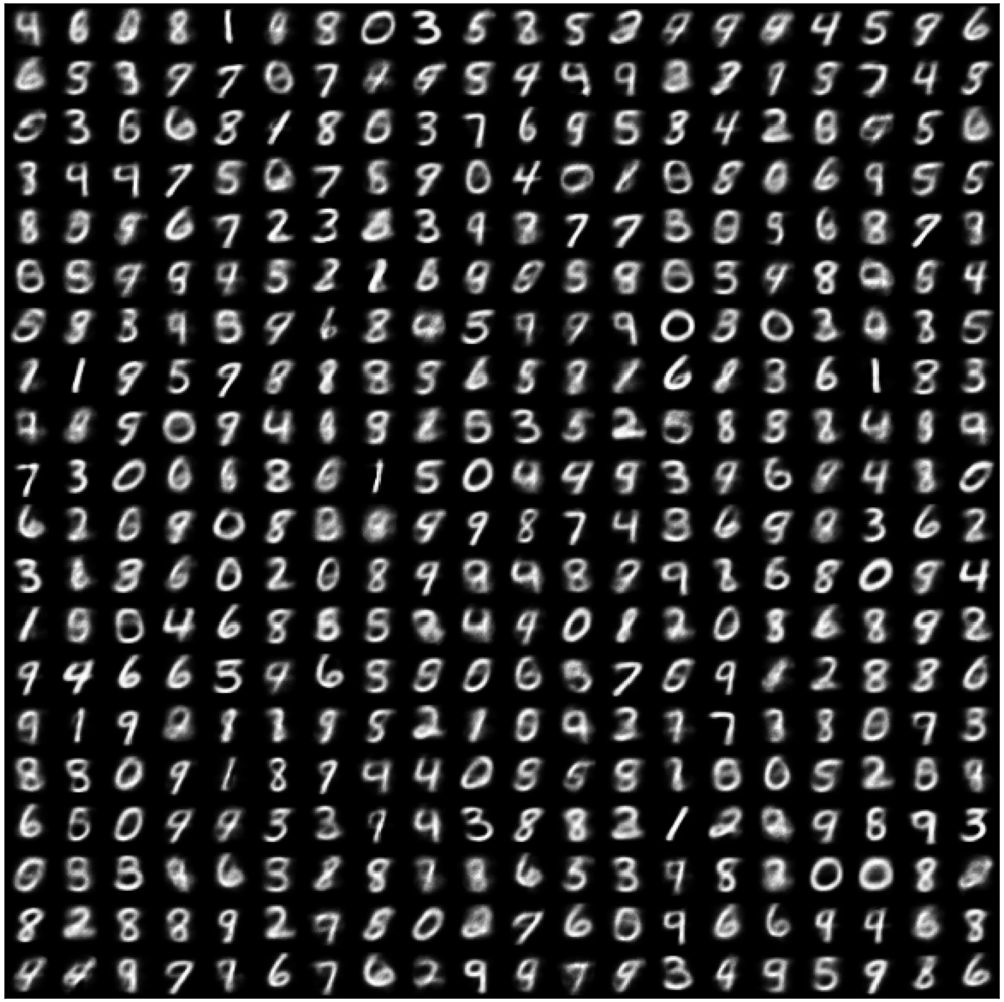
| Method | MNIST | TFD |
|---|---|---|
| DBN (Bengio et al., 2013) | $138 \pm 2$ | $1909 \pm 66$ |
| SCAE (Bengio et al., 2013) | $121 \pm 1.6$ | $2110 \pm 50$ |
| DGSN (Bengio et al., 2014) | $214 \pm 1.1$ | $1890 \pm 29$ |
| GAN (Goodfellow et al., 2014) | $225 \pm 2$ | $2057 \pm 26$ |
| GMMN (Li et al., 2015) | $147 \pm 2$ | $2085 \pm 25$ |
| IMLE (Proposed Method) | $\mathbf{257 \pm 6}$ | $\mathbf{2139 \pm 27}$ |

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

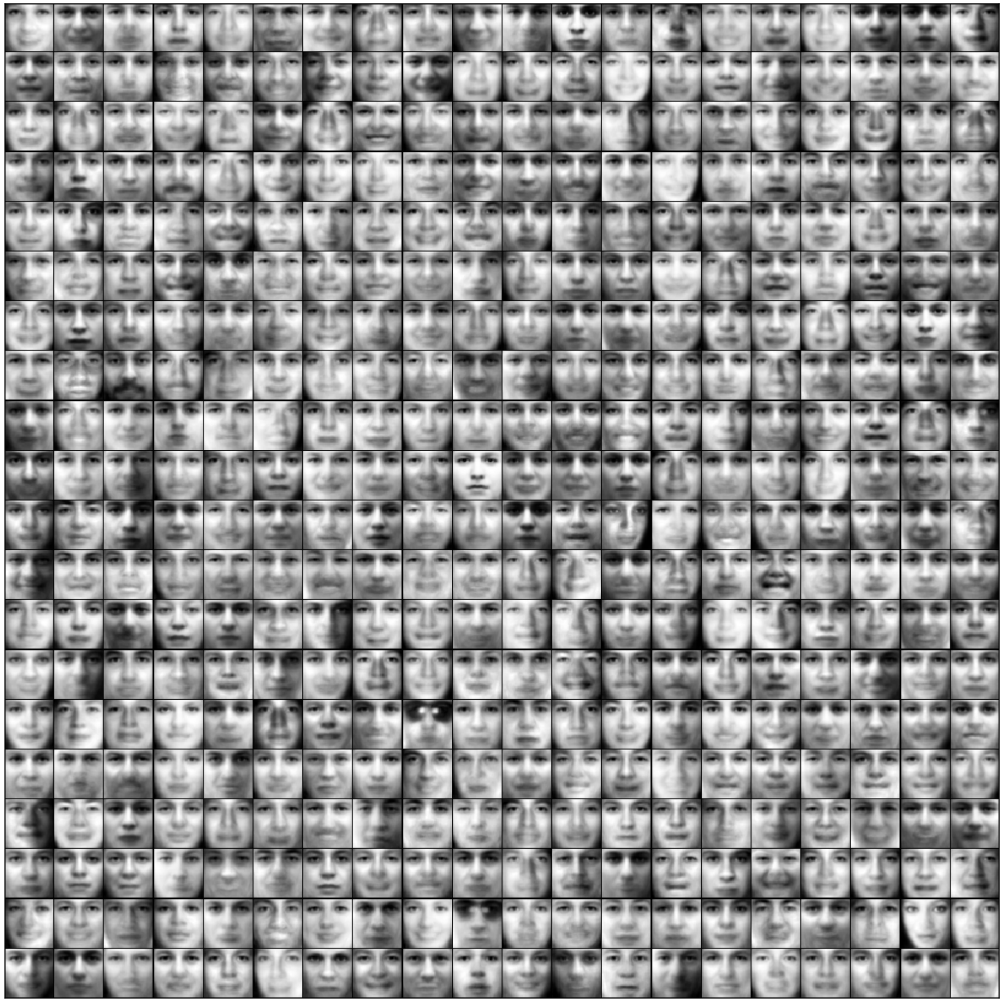
# Samples (MNIST)



K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

Berkeley
UNIVERSITY OF CALIFORNIA

# Samples (TFD)

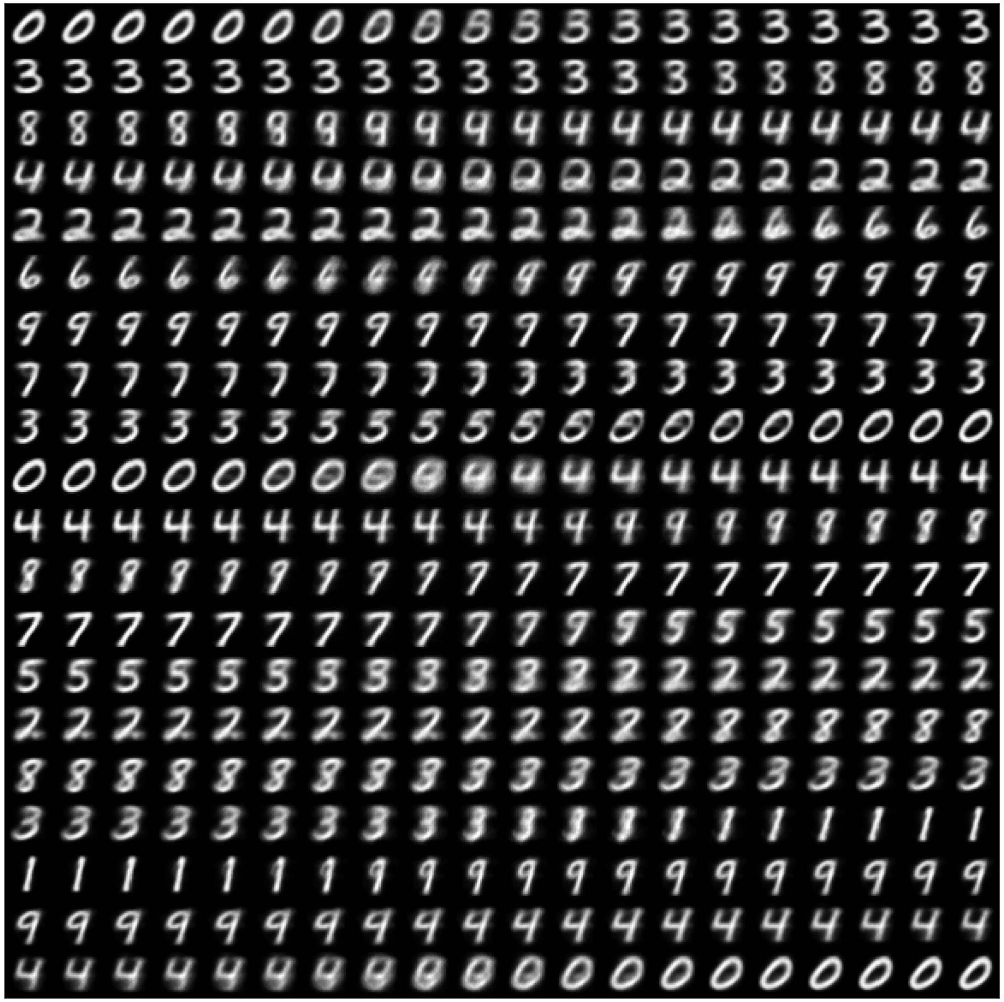


K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Samples (CIFAR-10)

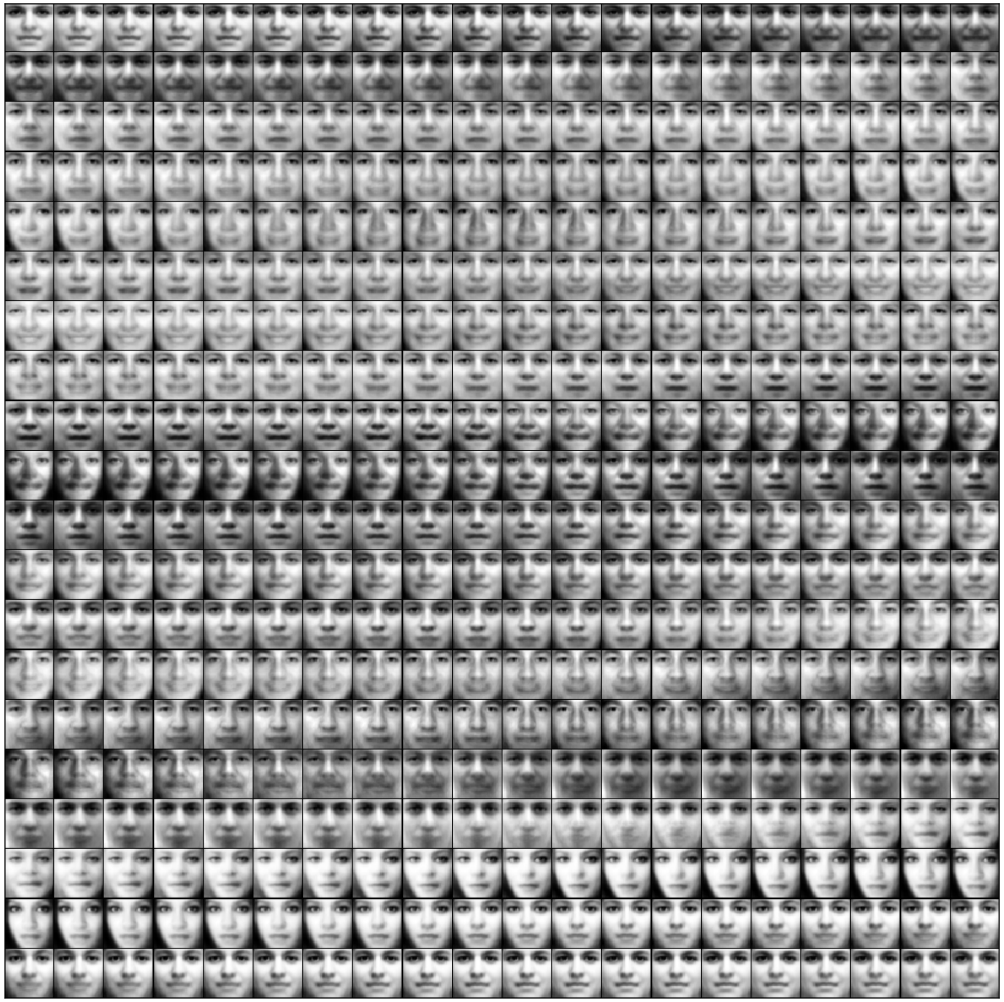
K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Interpolations (MNIST)



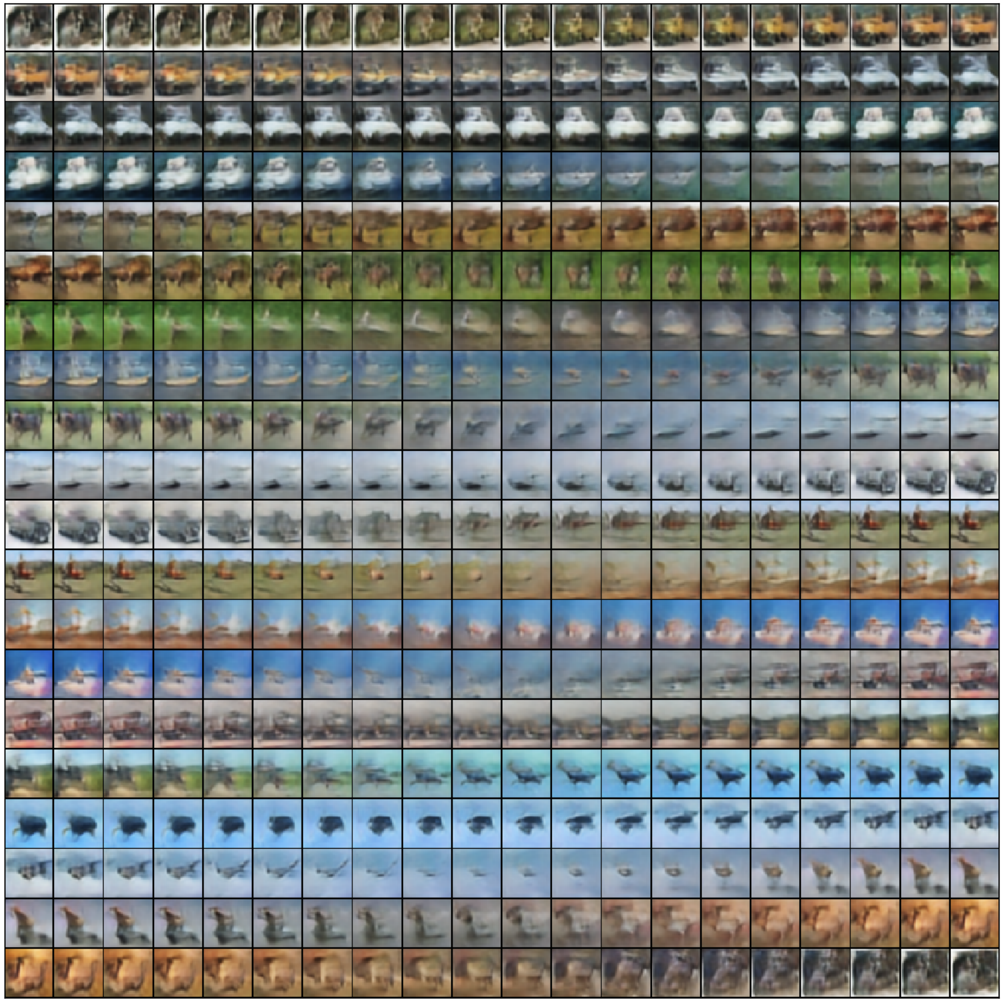
K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Interpolations (TFD)



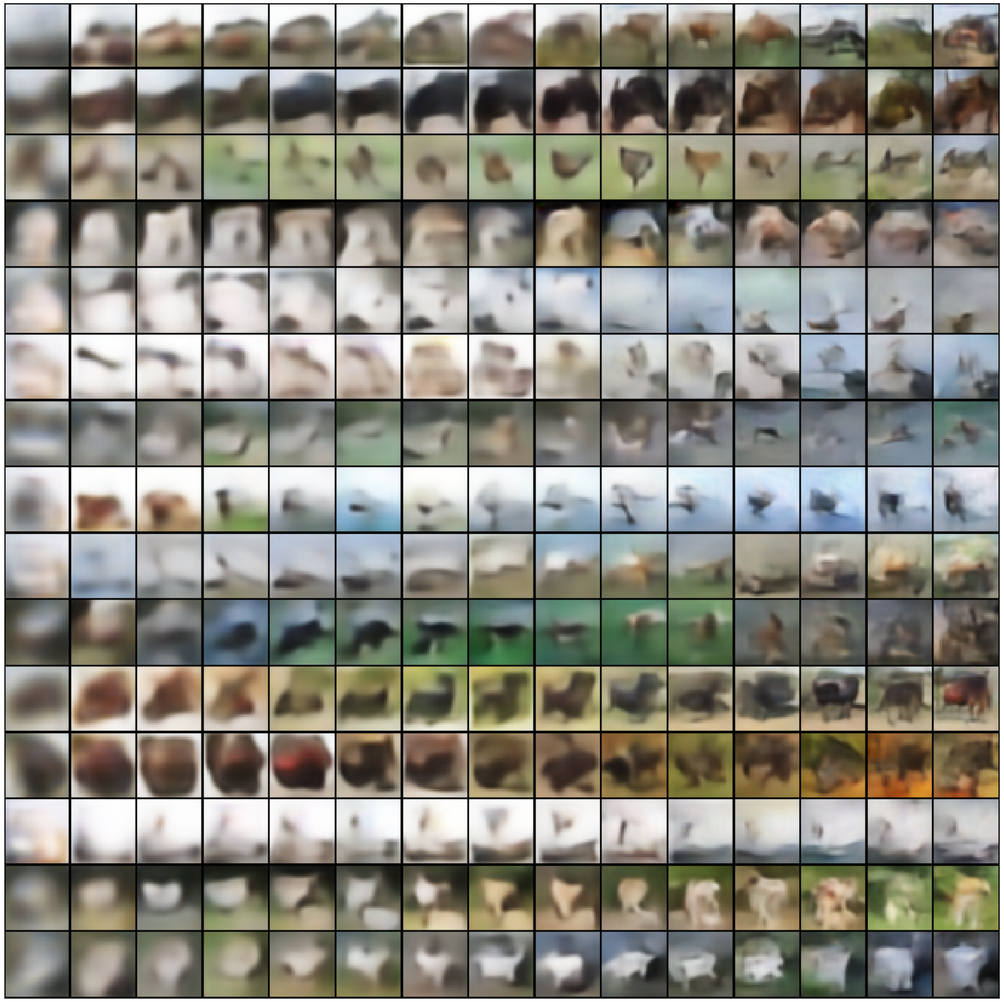
K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Interpolations (CIFAR-10)



K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

# Training Progress (CIFAR-10)



K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018

Berkeley
UNIVERSITY OF CALIFORNIA

# Agenda

1. Implicit Maximum Likelihood Estimation
2. Comparison to Generative Adversarial Nets (GANs)
3. Why Maximum Likelihood
4. Equivalence to Maximum Likelihood
5. *Fast Nearest Neighbour Search*
6. Applications to Conditional Image Synthesis

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



Berkeley
UNIVERSITY OF CALIFORNIA

# Fast Nearest Neighbour Search

Berkeley
UNIVERSITY OF CALIFORNIA

# Definition

- Given a database of $n$ points and the query, find the $k$ points that are closest to the query.

K. Li and J. Malik. Fast $k$-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Notions of Dimensionality

- The hardness of a dataset can be characterized using two notions of dimensionality.
  - Ambient dimensionality: the dimensionality of the space that contains the data points.
  - Intrinsic dimensionality: can be roughly thought of as the dimensionality of the data manifold.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Intrinsic Dimensionality

- *Definition:*

  A dataset $D$ has intrinsic dimensionality[1] $d'$ if for all $r > 0$, $\alpha > 1$ and $p$ such that $|B_p(r)| \geq k$,

$$|B_p(\alpha r)| \leq \alpha^{d'} |B_p(r)|$$

[1]This is also known as the expansion dimension or the KR-dimension.

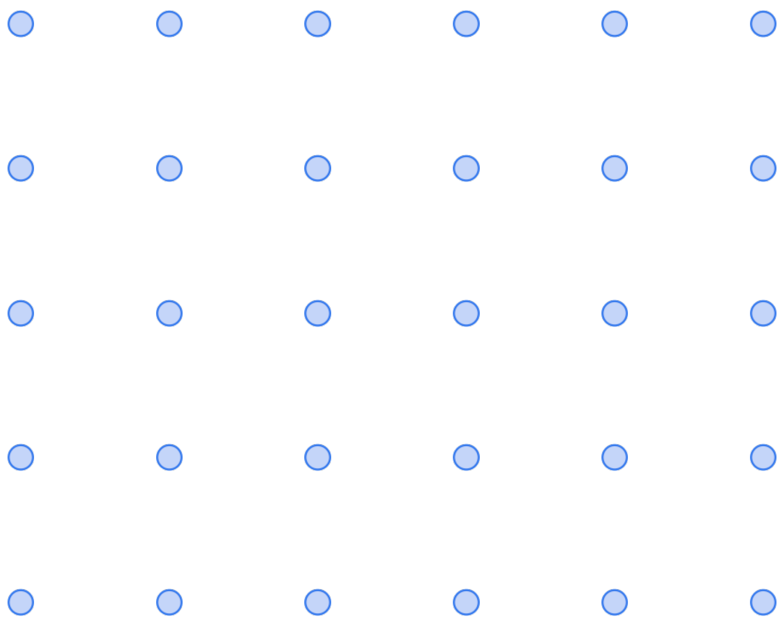
K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Intrinsic Dimensionality

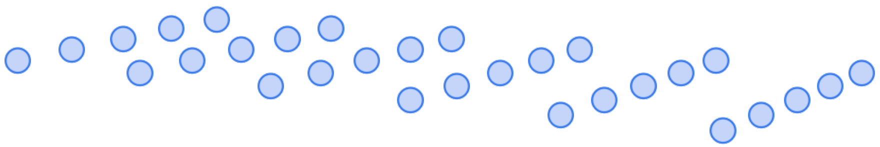
- A $d'$-dimensional uniform grid $\mathbb{Z}^{d'}$ has intrinsic dimensionality $d'$.



K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# Intrinsic Dimensionality

- If it were embedded in a higher-dimensional space, it would retain its intrinsic dimensionality.



K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Why is High Dimensionality Hard?

$$d' = 1$$

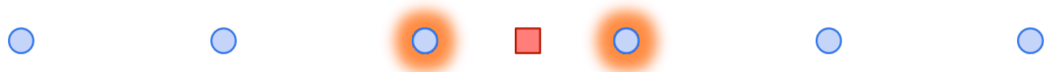
K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Why is High Dimensionality Hard?

$$d' = 1$$



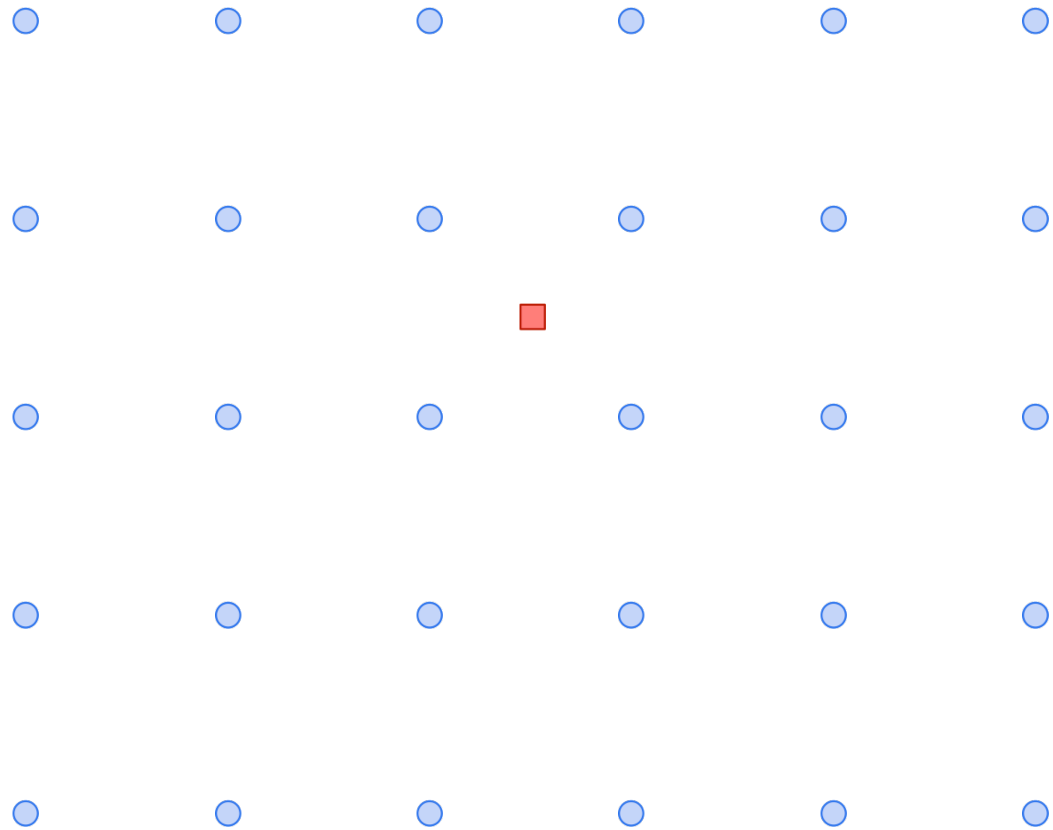
K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Why is High Dimensionality Hard?

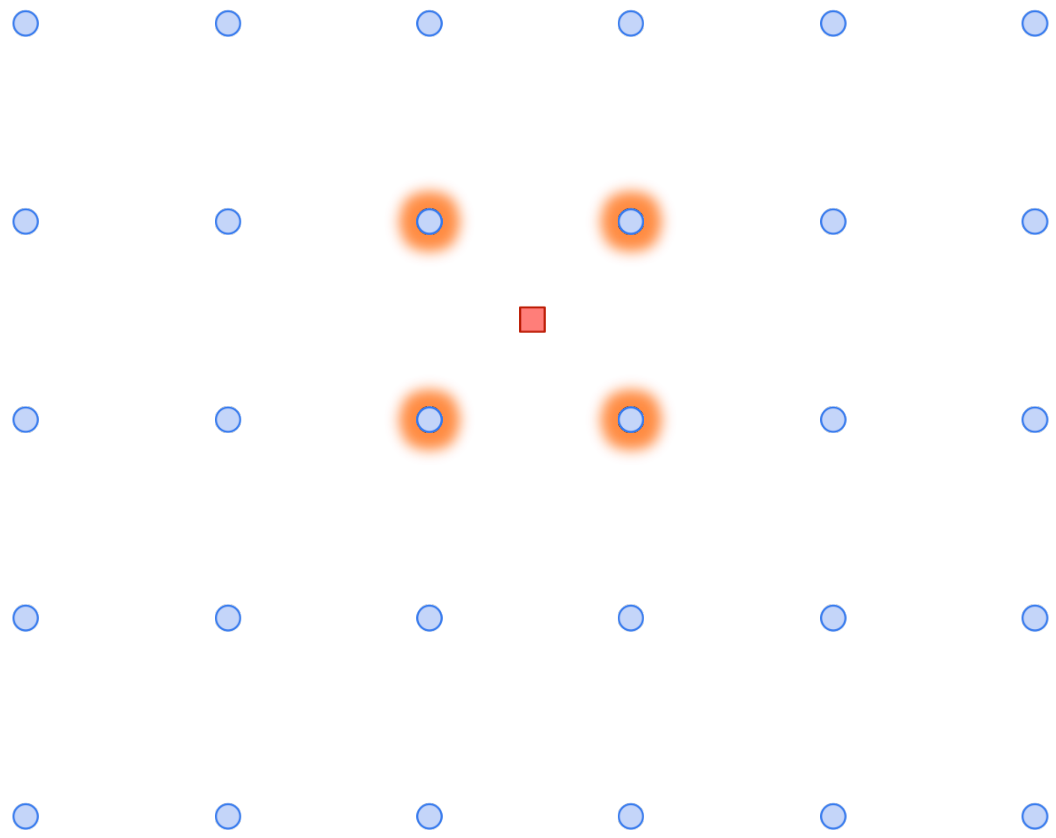
$d' = 2$



K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# Why is High Dimensionality Hard?

$$d' = 2$$



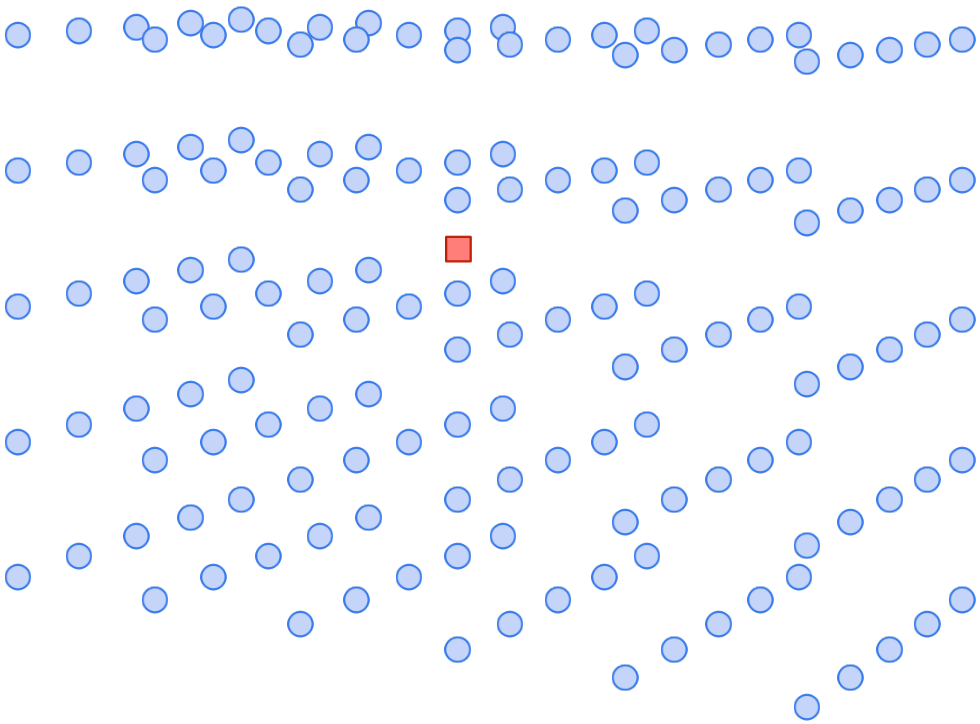
K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Why is High Dimensionality Hard?

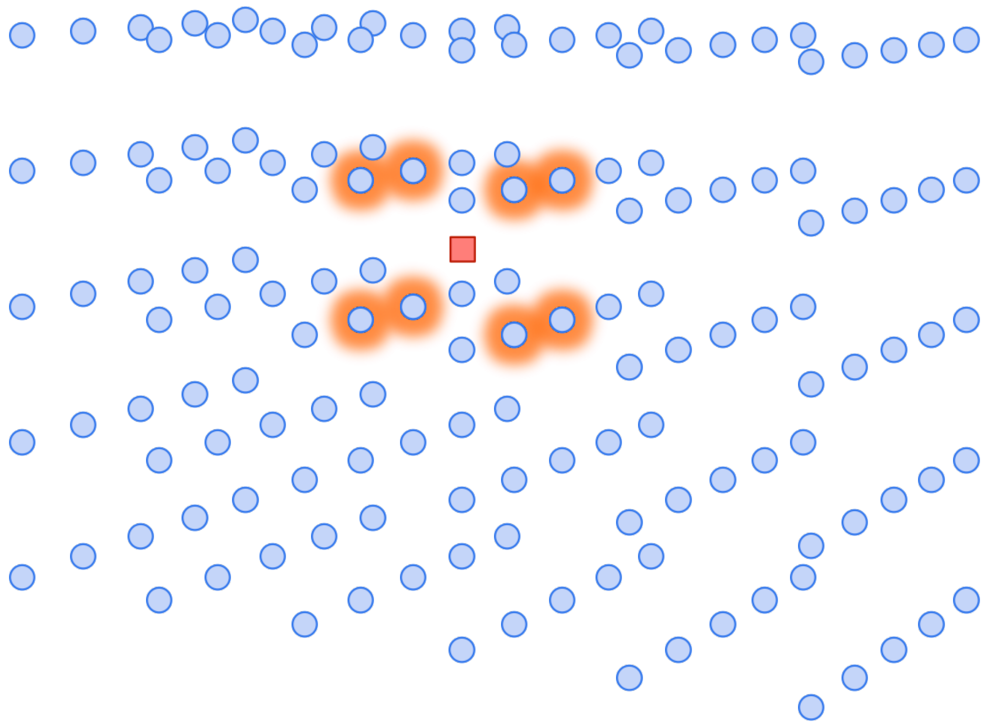
$$d' = 3$$



K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# Why is High Dimensionality Hard?

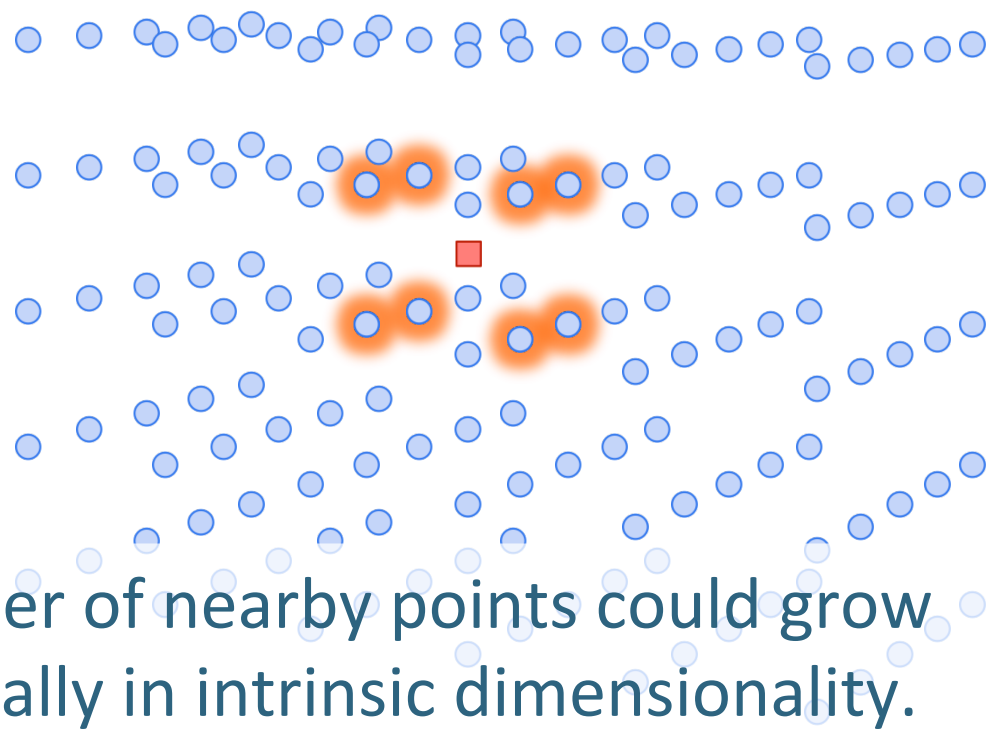
$$d' = 3$$



K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# Why is High Dimensionality Hard?

$$d' = 3$$



- The number of nearby points could grow exponentially in intrinsic dimensionality.

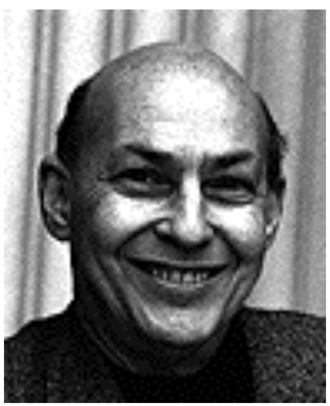
K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

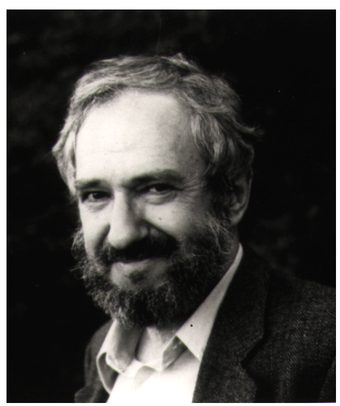


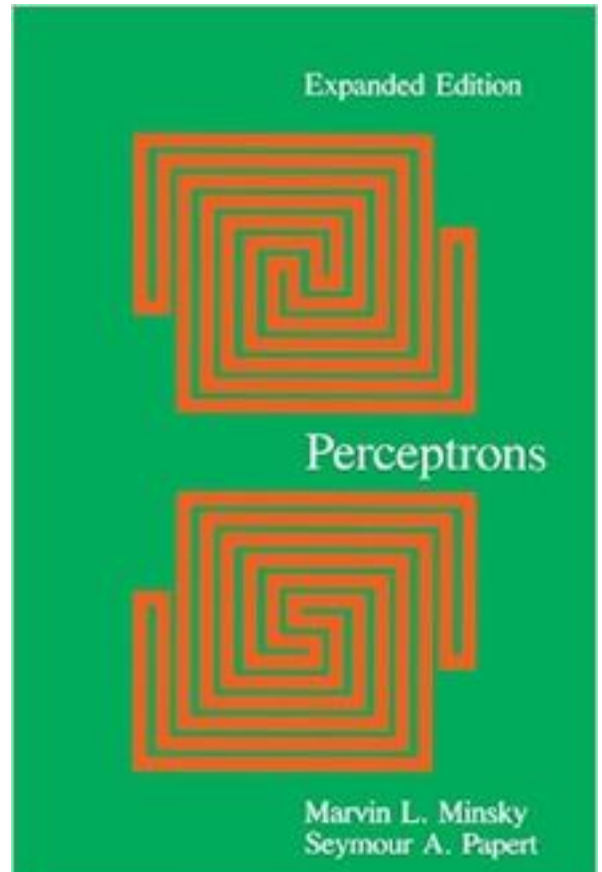
Berkeley
UNIVERSITY OF CALIFORNIA

# History

- The problem of nearest neighbour search was posed by Minsky and Papert in their seminal book, *Perceptrons* (1969).



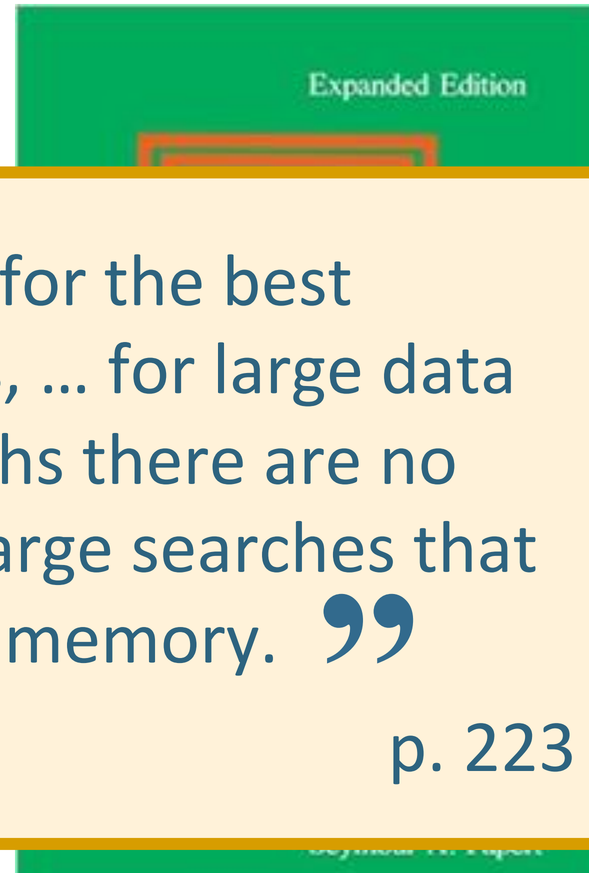





K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# History

- The problem of nearest neighbour search was




“We conjecture that even for the best possible $\mathbf{A}_{\mathrm{file}}$ - $\mathbf{A}_{\mathrm{find}}$ pairs, … for large data sets with long word lengths there are no practical alternatives to large searches that inspect large parts of the memory.”

p. 223

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# History

- The problem of nearest neighbour search was

“ We conjecture that even for the best possible $\mathbf{A}_{\text{file}}$ - $\mathbf{A}_{\text{find}}$ pairs, ... for large data sets with long word lengths there are no practical alternatives to large searches that inspect large parts of the memory. ”

p. 223

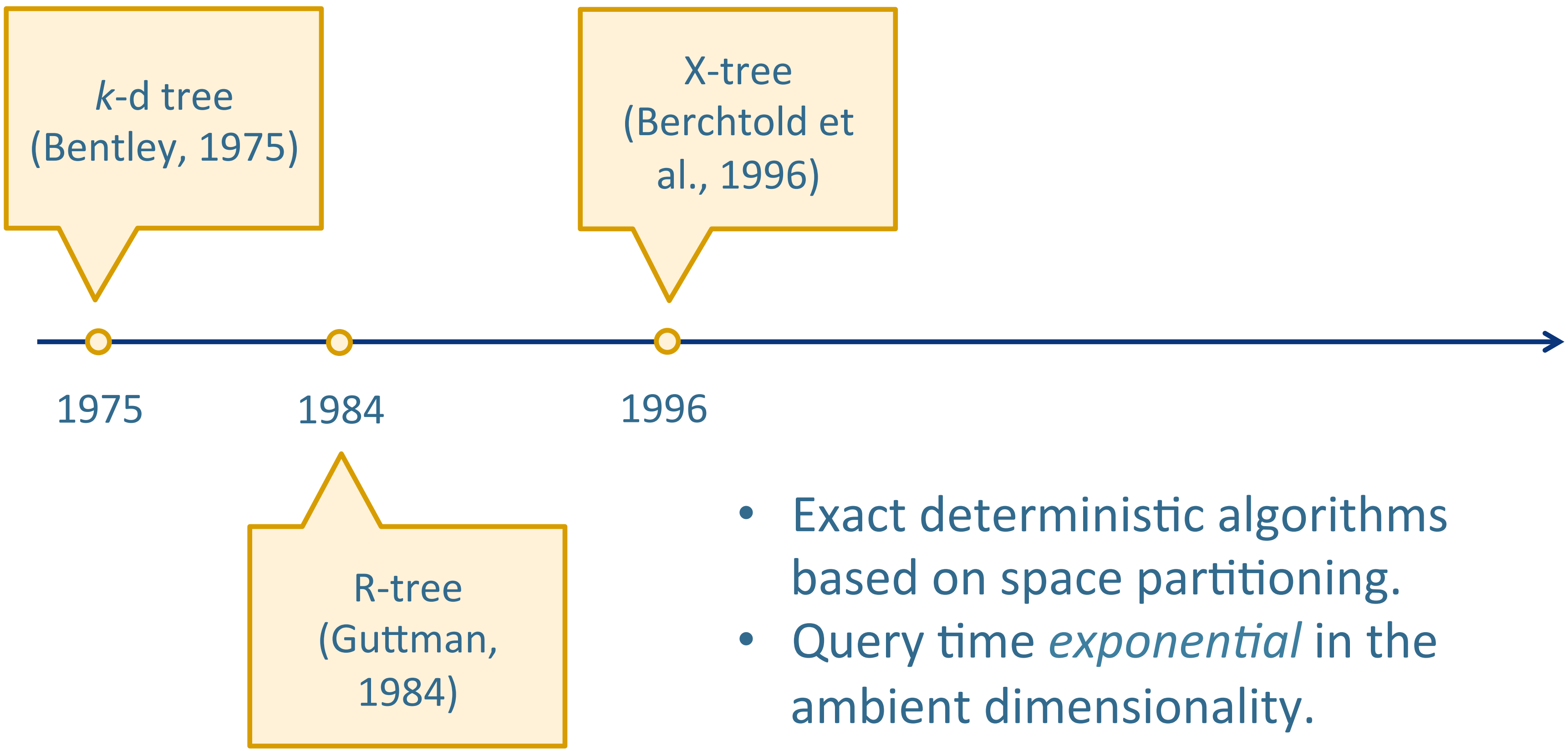
In other words, even for the best choice of dataset and queries, when the dataset is large and high-dimensional, doing substantially better than exhaustive search is conjectured to be impossible.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# The Curse of Dimensionality



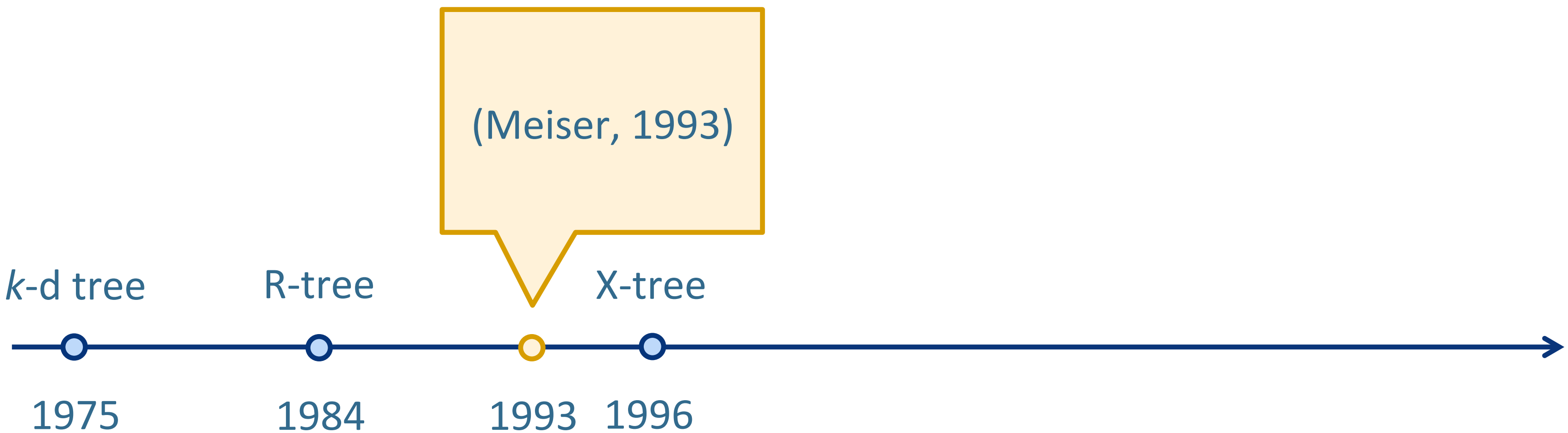

- Exact deterministic algorithms based on space partitioning.
- Query time *exponential* in the ambient dimensionality.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# The Curse of Dimensionality



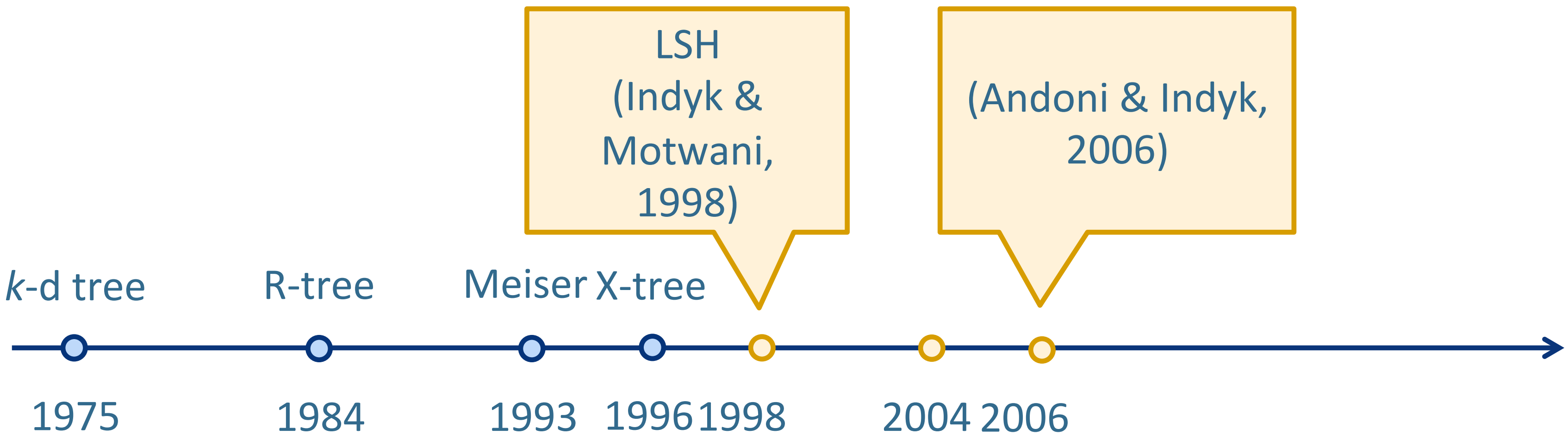

- Exact deterministic algorithm.
- Query time *polynomial* in ambient dimensionality.
- Space complexity *exponential* in ambient dimensionality.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley UNIVERSITY OF CALIFORNIA

# The Curse of Dimensionality





- LSH introduced the idea of randomization.
- Approximate randomized algo-rithm based on space partitioning.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# The Curse of Dimensionality

LSH (Indyk & Motwani, 1998)

(Andoni & Indyk, 2006)

*k*-d tree — 1975

R-tree — 1984

Meiser — 1993

X-tree — 1996

1998

2004

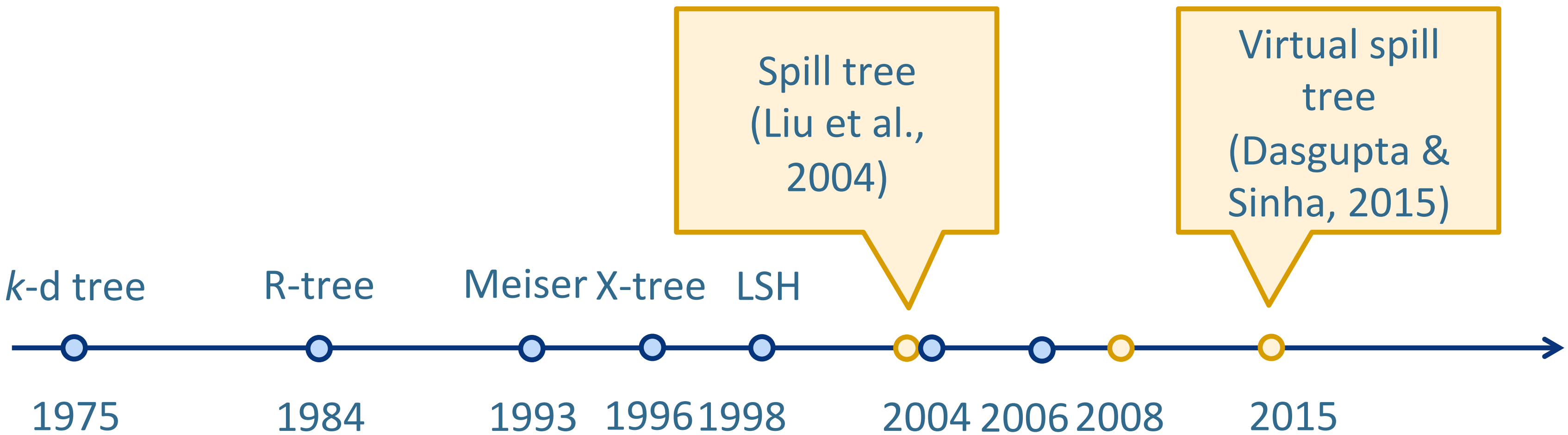
2006

Euclidean LSH (Datar et al., 2004)

- Query time is $O(dn^{\rho})$, where $\rho \approx 1/(1+\epsilon)^2$.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# The Curse of Dimensionality





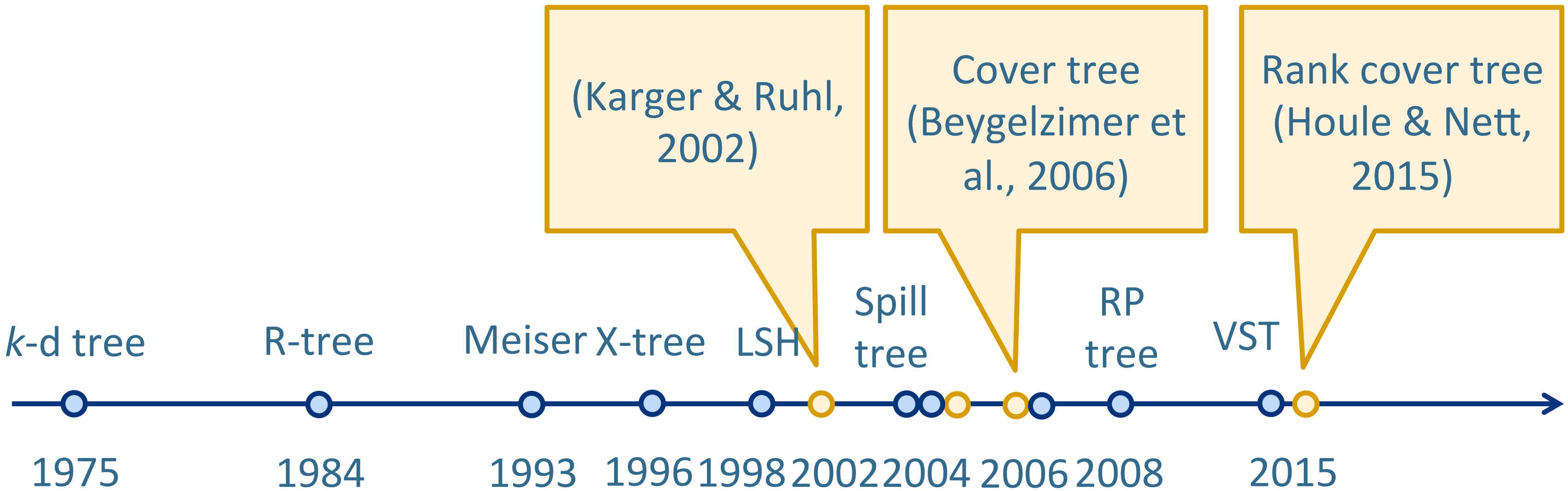
- Exact randomized algorithms based on space partitioning.
- Query time *exponential* in intrinsic dimensionality.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley UNIVERSITY OF CALIFORNIA

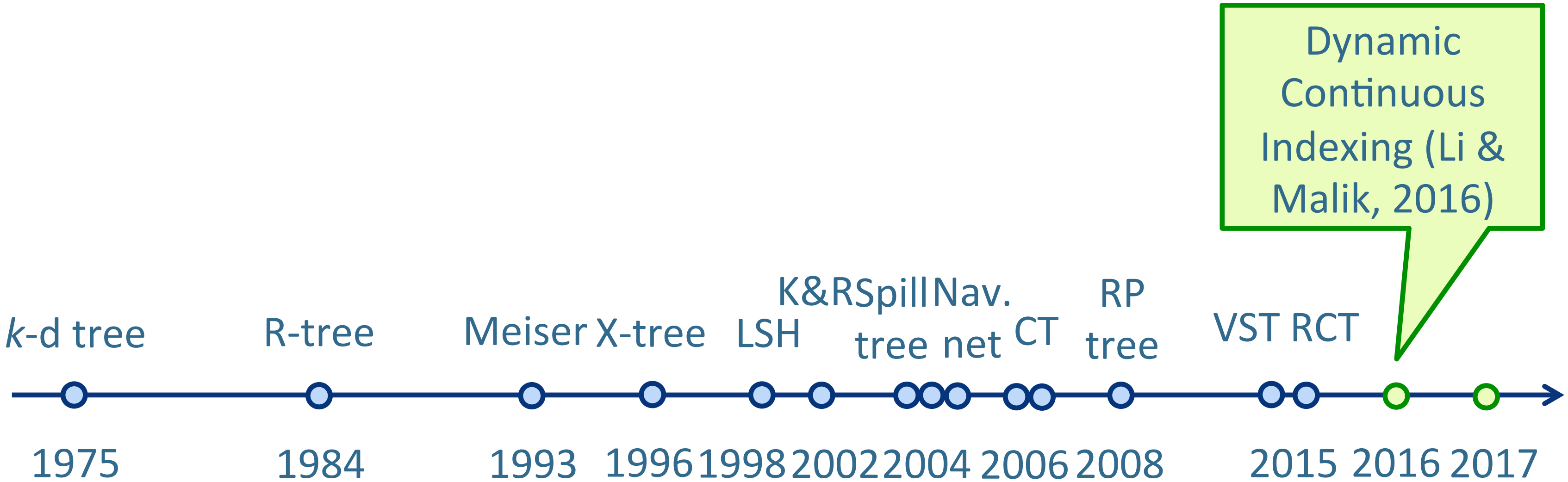
# The Curse of Dimensionality




- Exact algorithms based on local search and coarse-to-fine.
- Query time *exponential* in intrinsic dimensionality.


K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# The Curse of Dimensionality




- Our contribution: a new family of exact randomized algorithms, known as Dynamic Continuous Indexing.

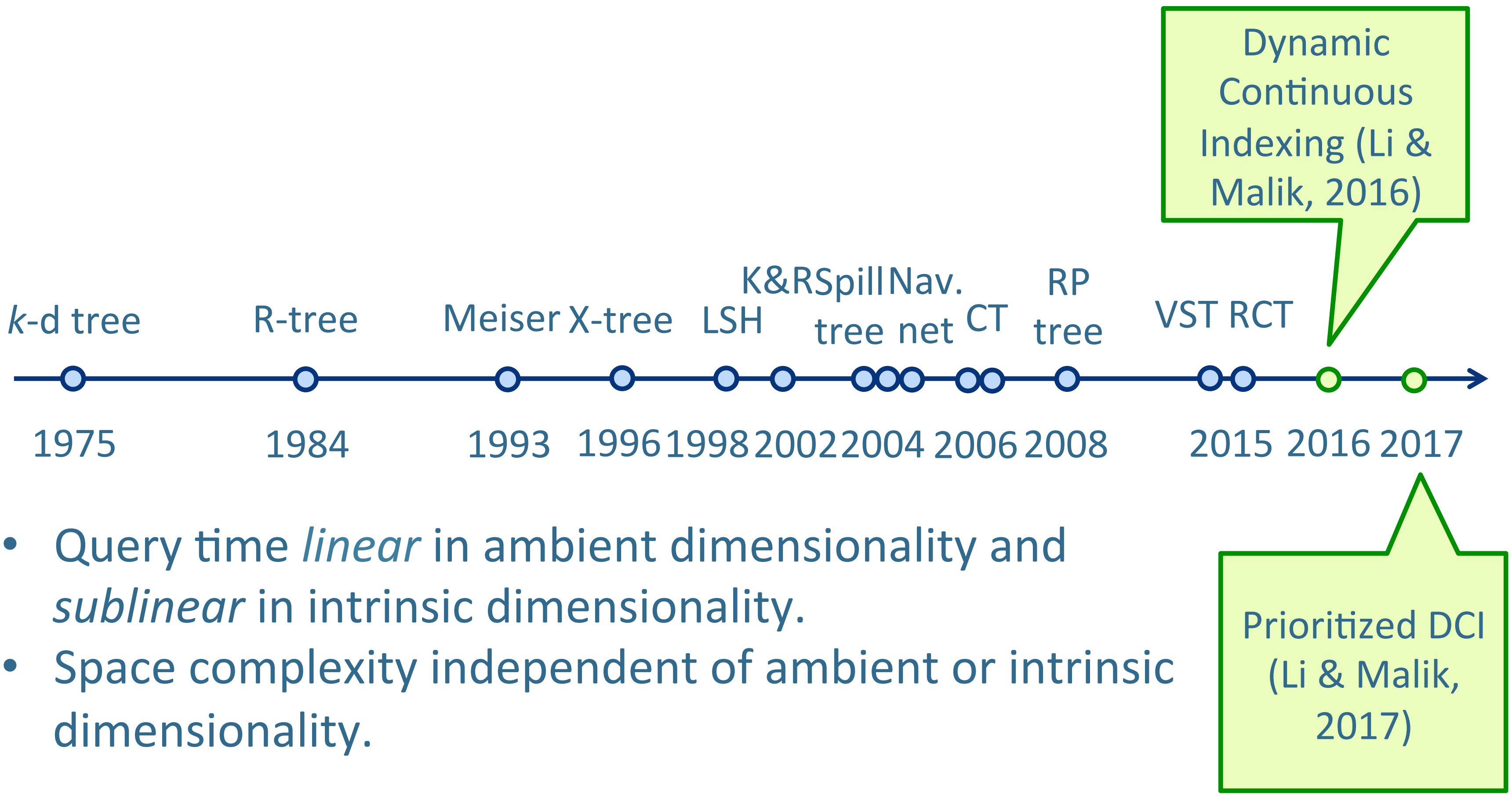
Prioritized DCI (Li & Malik, 2017)

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley UNIVERSITY OF CALIFORNIA

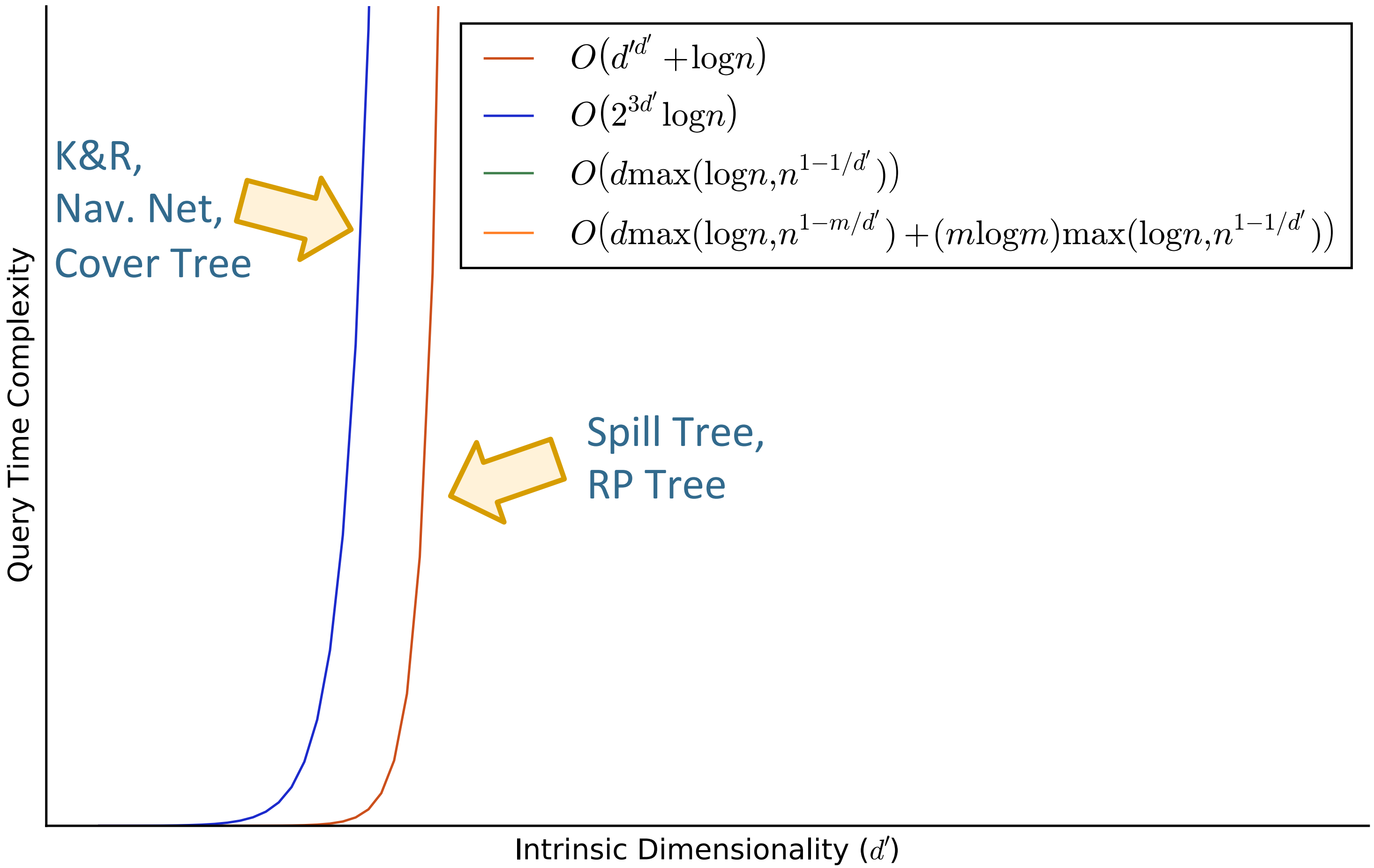
# The Curse of Dimensionality




- Query time *linear* in ambient dimensionality and *sublinear* in intrinsic dimensionality.
- Space complexity independent of ambient or intrinsic dimensionality.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# The Curse of Dimensionality

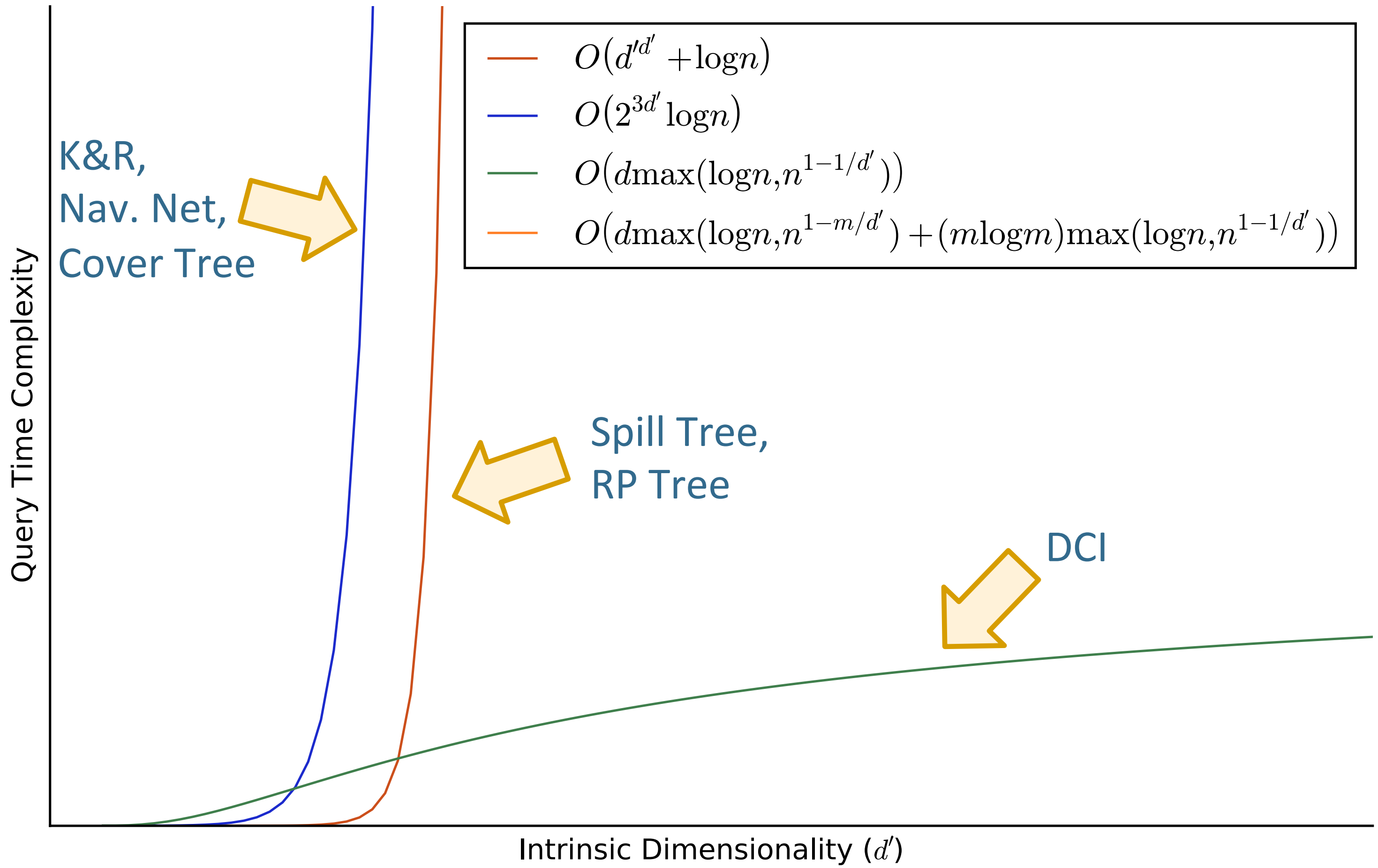



K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



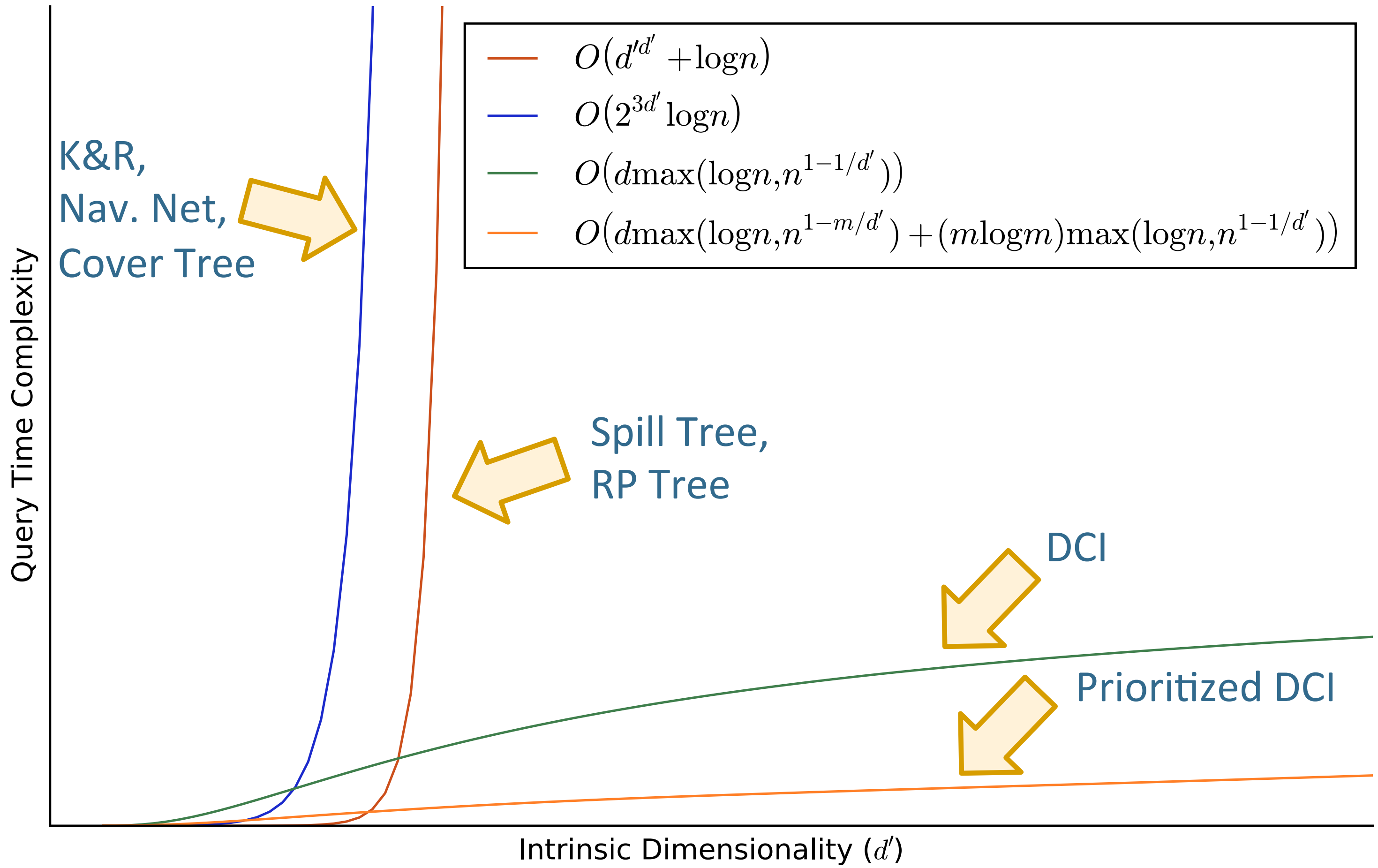
Berkeley
UNIVERSITY OF CALIFORNIA

# The Curse of Dimensionality




K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# The Curse of Dimensionality




K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Our Approach

- Key difference from prior methods: Dynamic Continuous Indexing (DCI) avoids *space partitioning*.
- Space partitioning is a divide-and-conquer strategy that underlies most existing methods, including *k*-d trees and locality-sensitive hashing (LSH).
  - It works by partitioning the space into discrete cells and keeping track of points contained in each.
- We conjecture that the curse of dimensionality stems from the inherent deficiencies of space partitioning.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



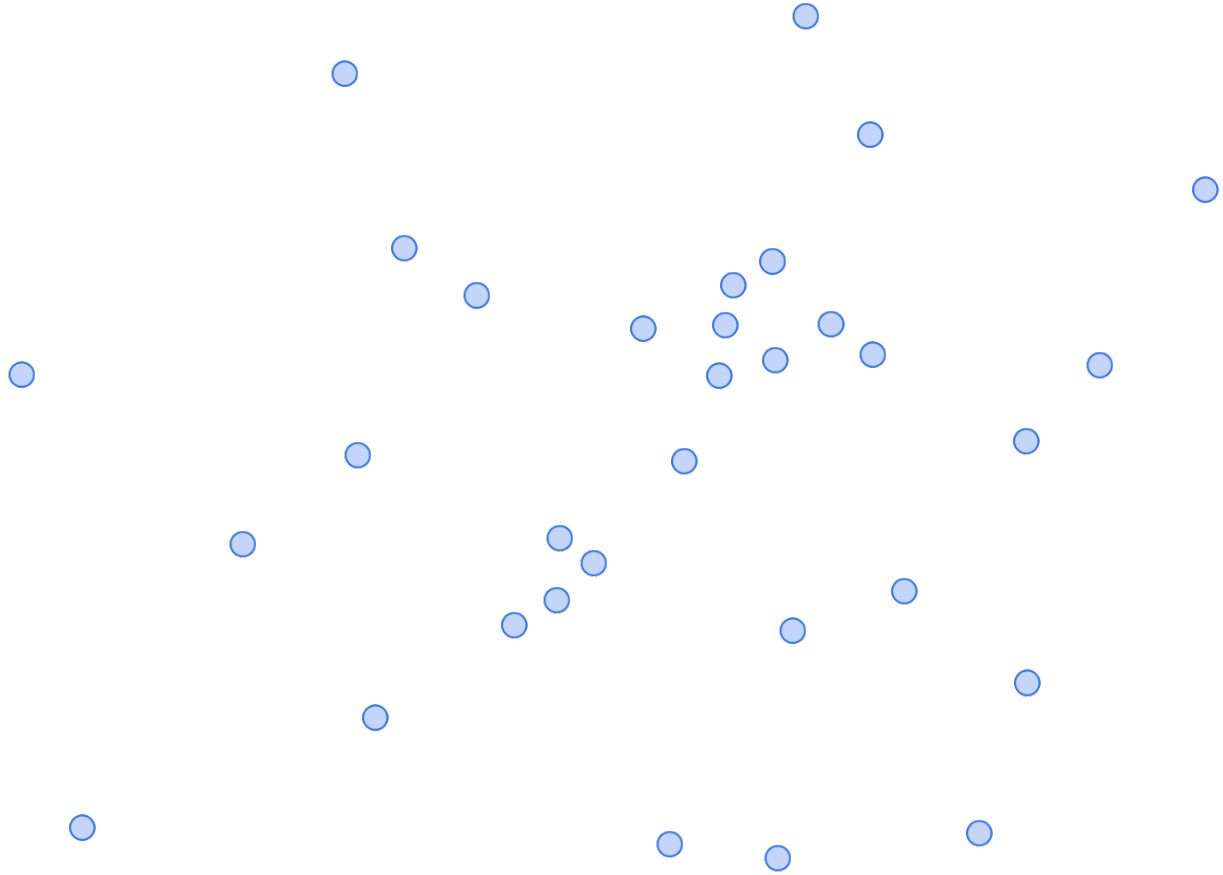
Berkeley
UNIVERSITY OF CALIFORNIA

# The Case Against Space Partitioning



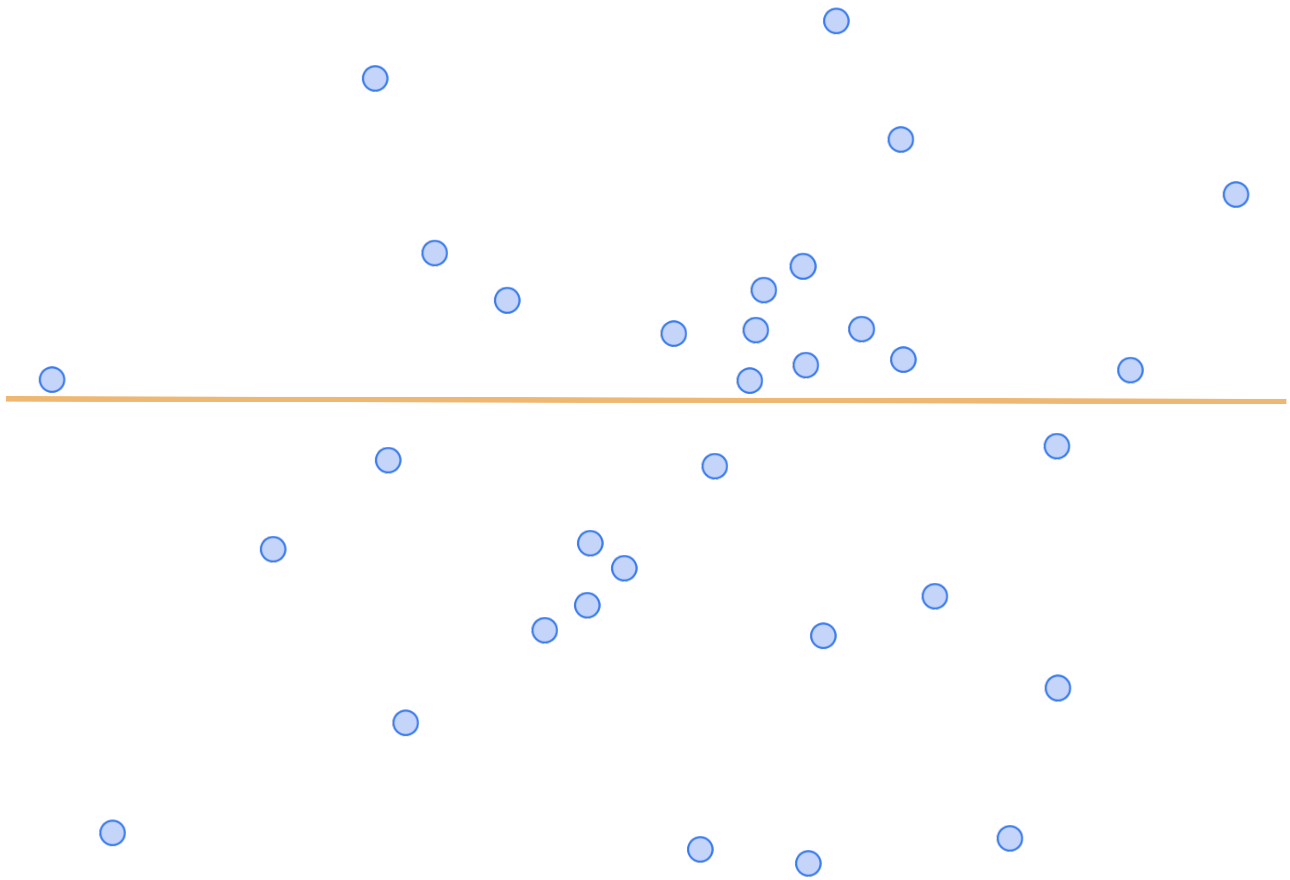
Berkeley
UNIVERSITY OF CALIFORNIA

# *k*-d tree



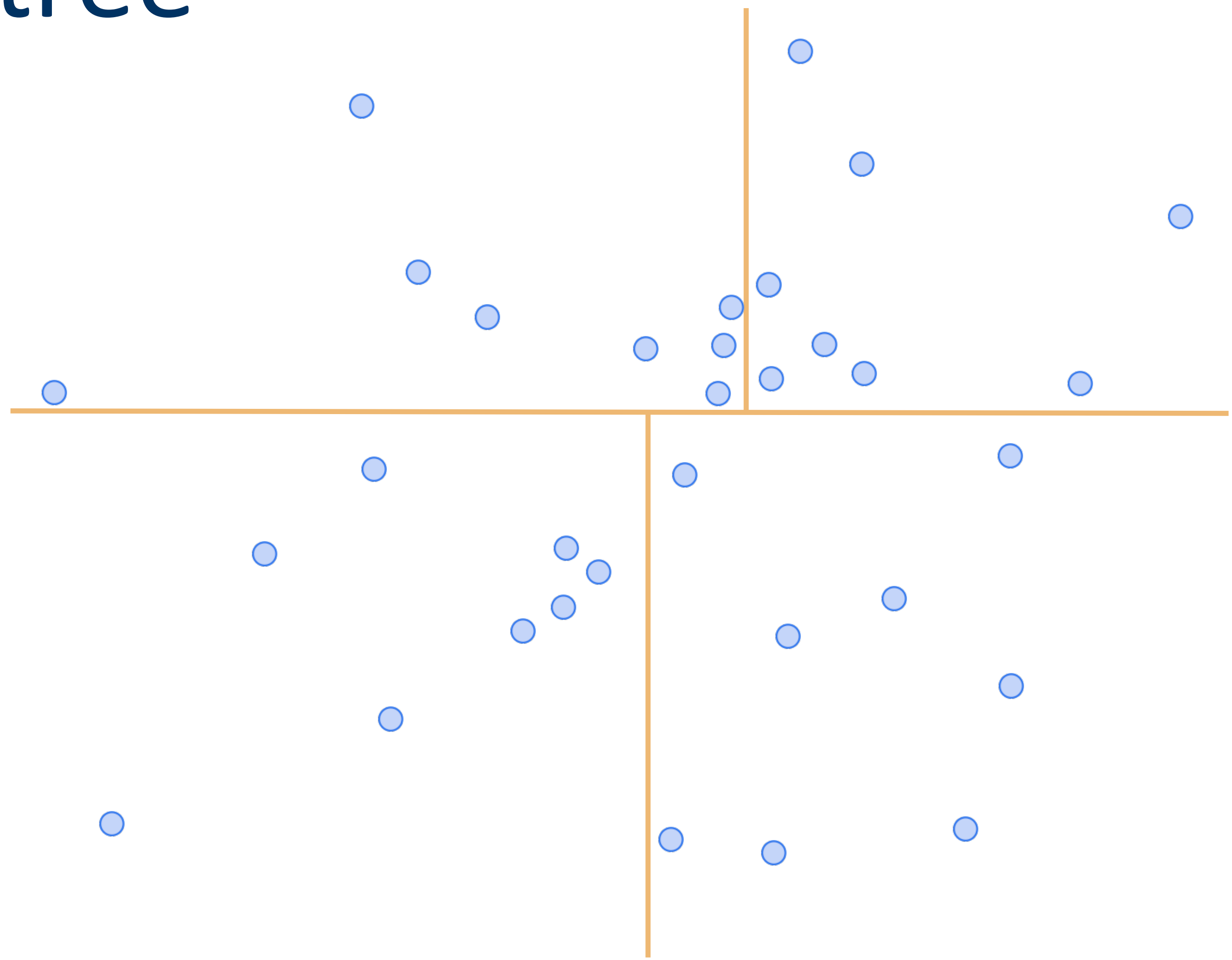
K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# *k*-d tree



K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# *k*-d tree

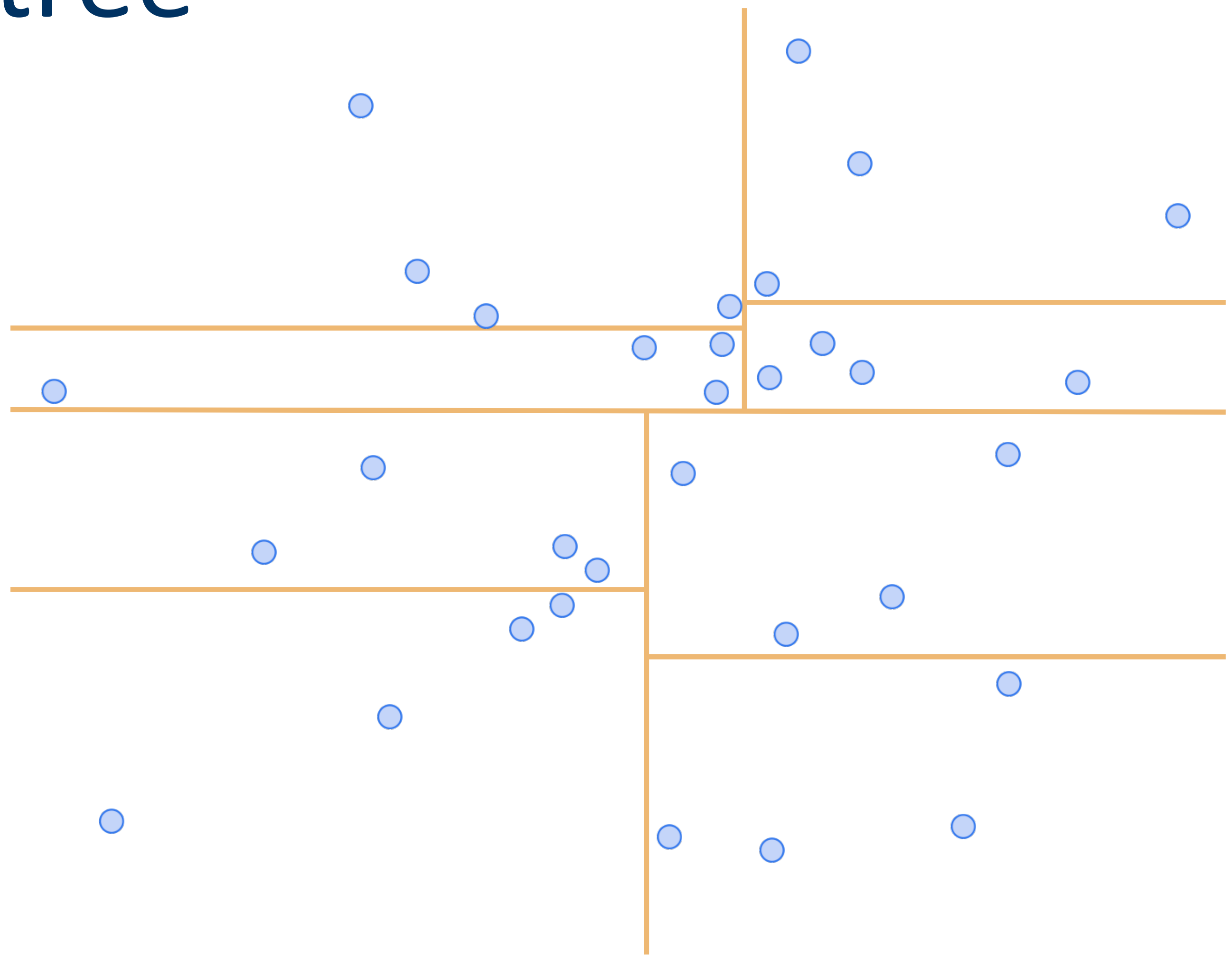


K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# *k*-d tree

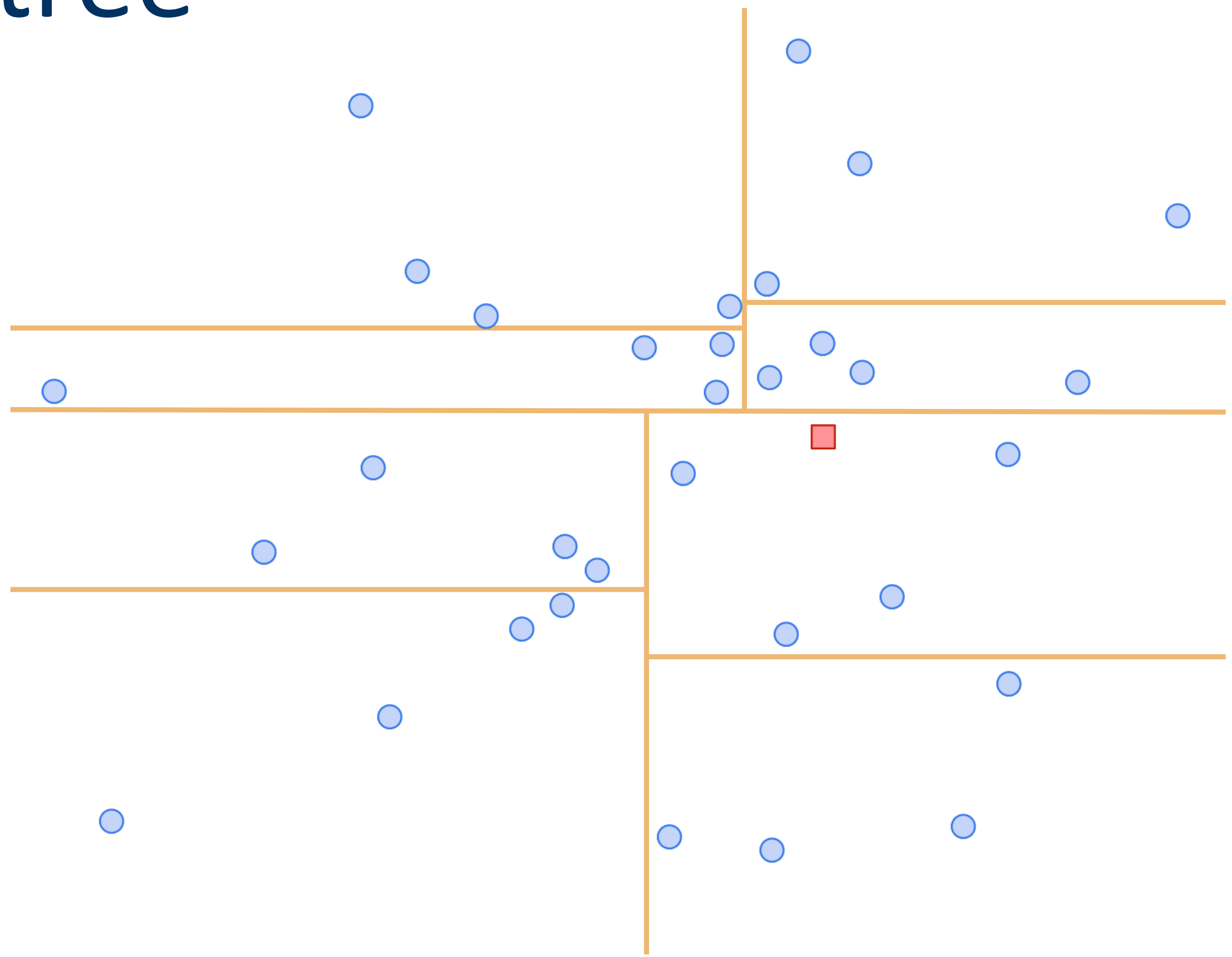


K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# *k*-d tree

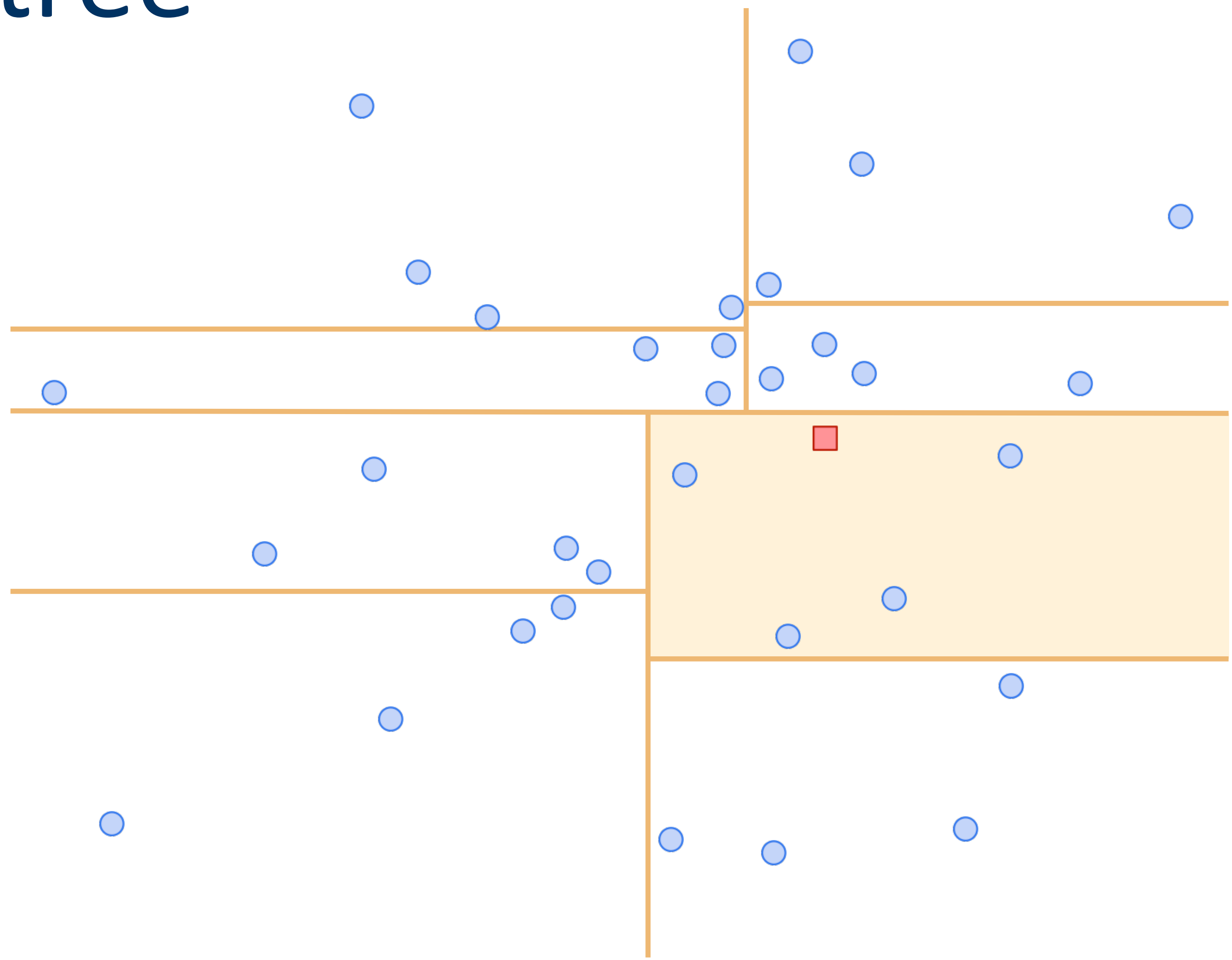


K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



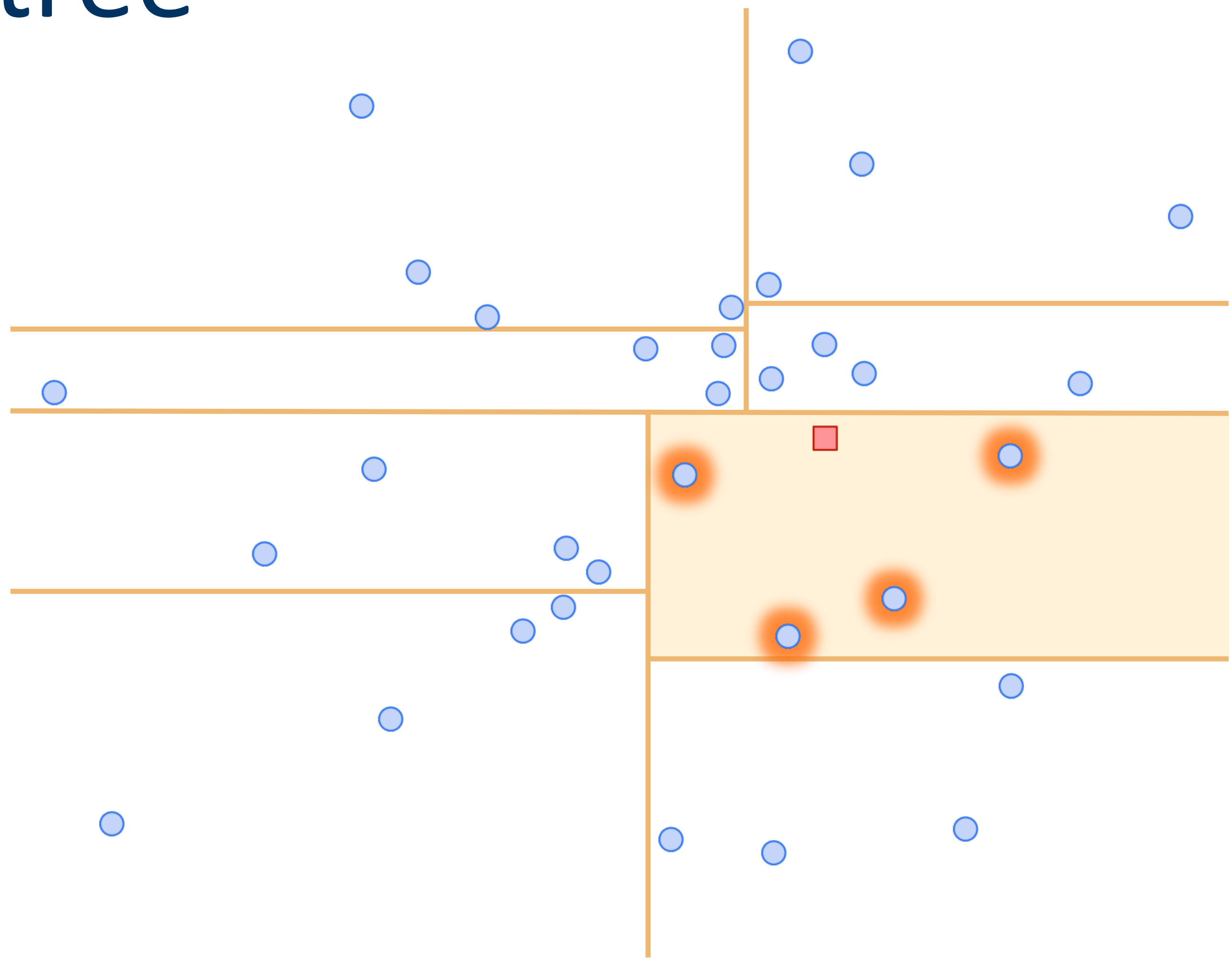
Berkeley
UNIVERSITY OF CALIFORNIA

# *k*-d tree



K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# *k*-d tree

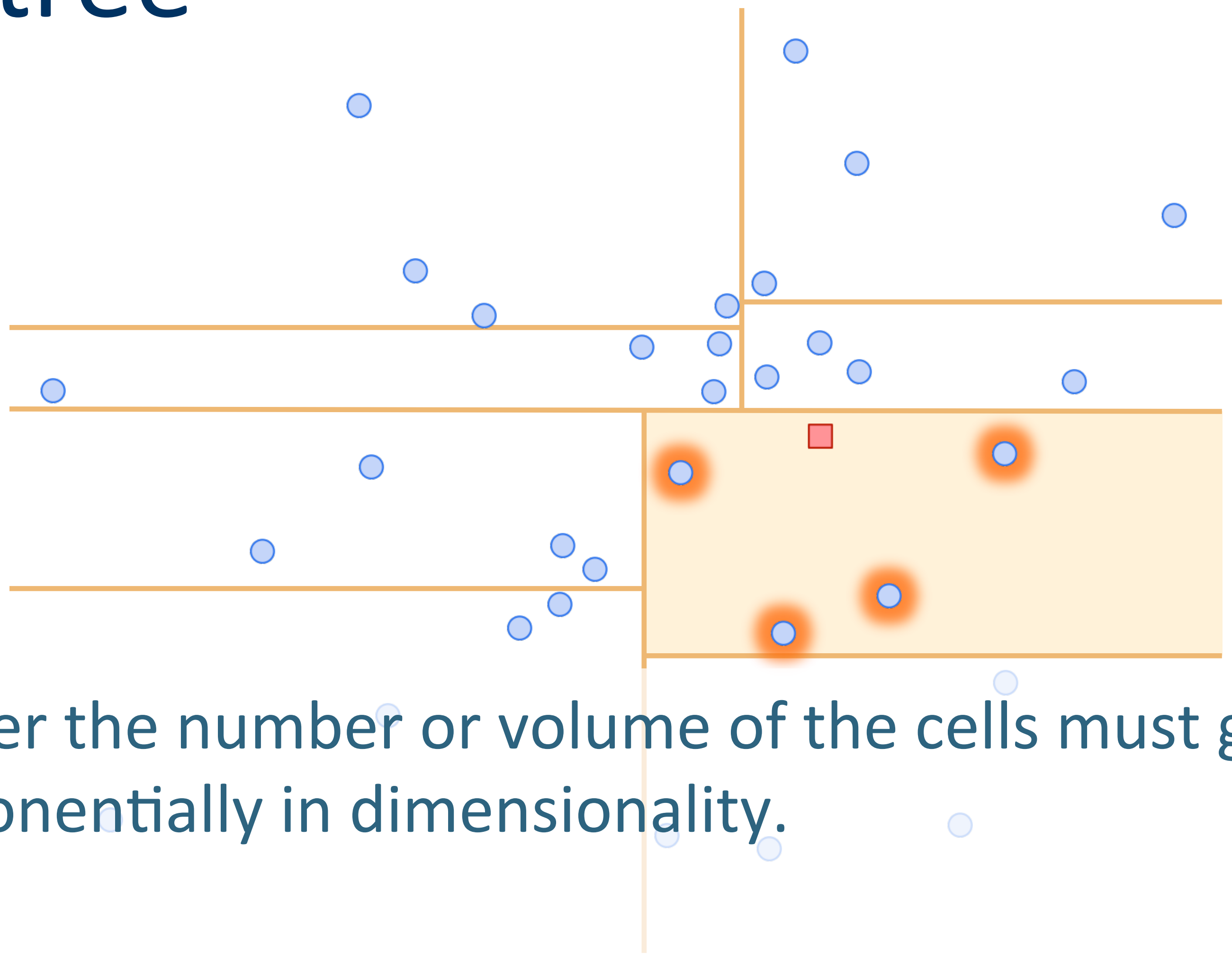


K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

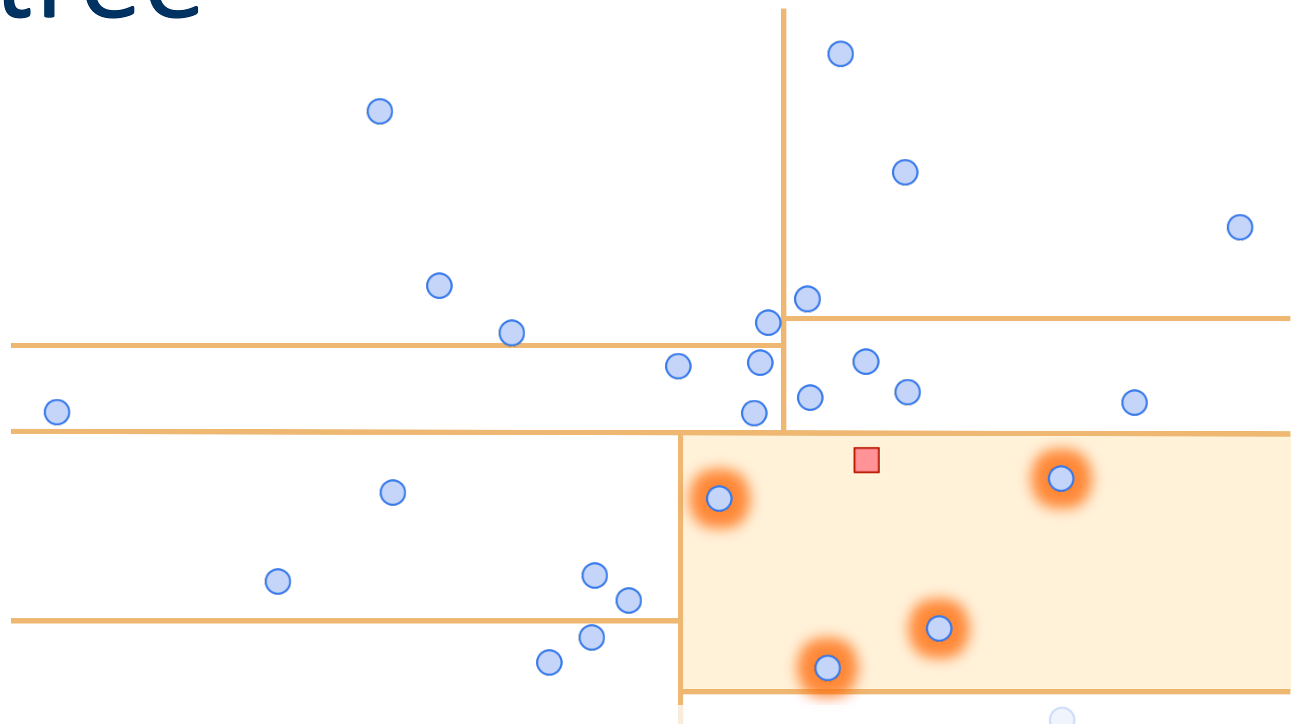
# *k*-d tree



- Either the number or volume of the cells must grow exponentially in dimensionality.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# *k*-d tree



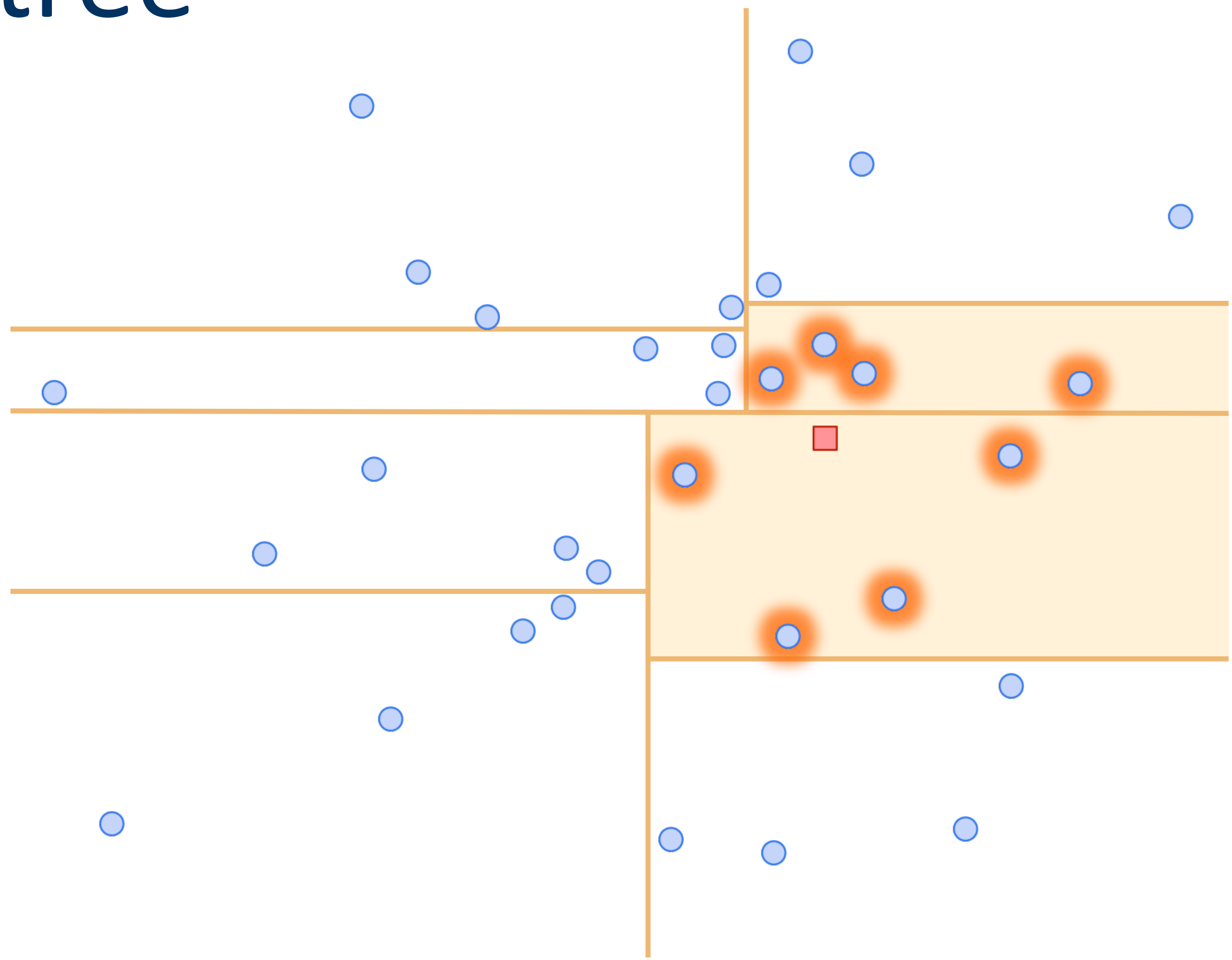
- Either the number or volume of the cells must grow exponentially in dimensionality.
- “Field of view” limited to cell containing the query.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



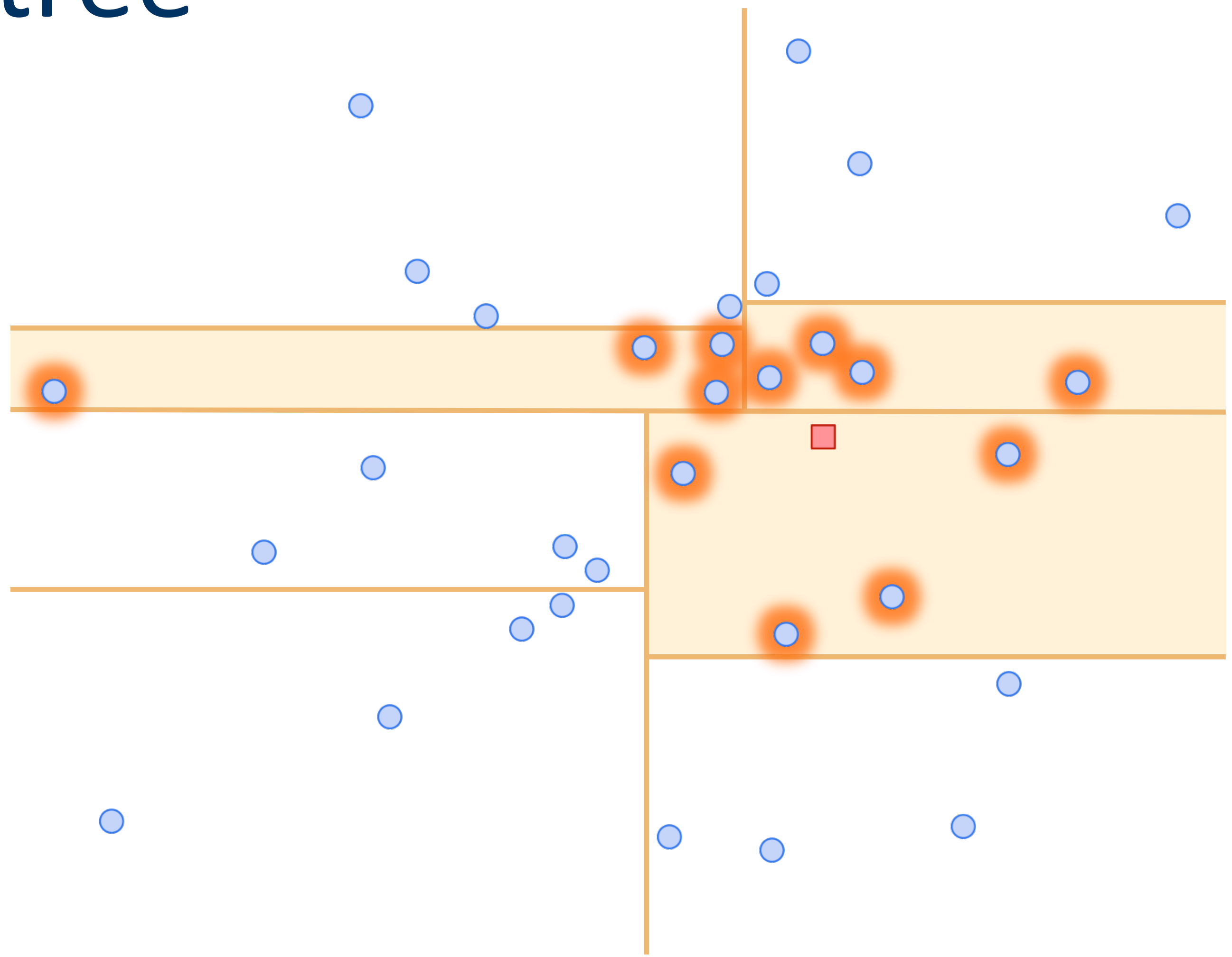
Berkeley
UNIVERSITY OF CALIFORNIA

# *k*-d tree



K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# *k*-d tree

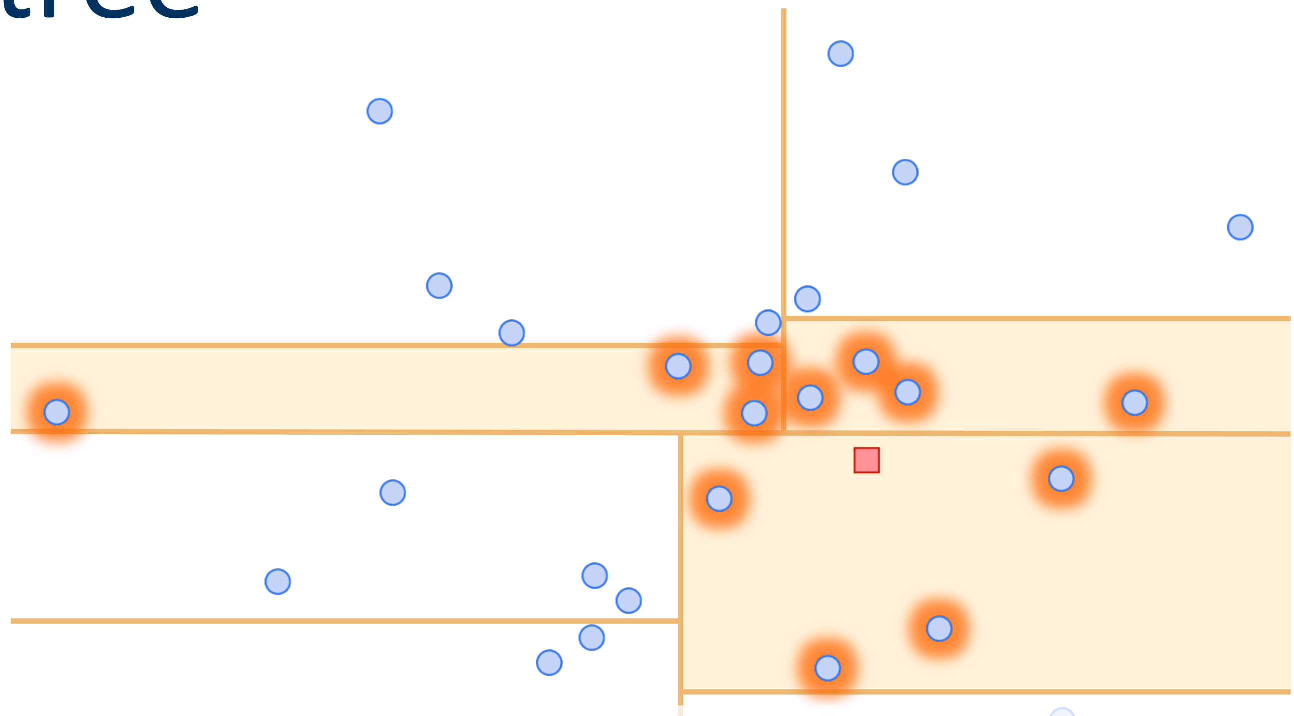


K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# *k*-d tree



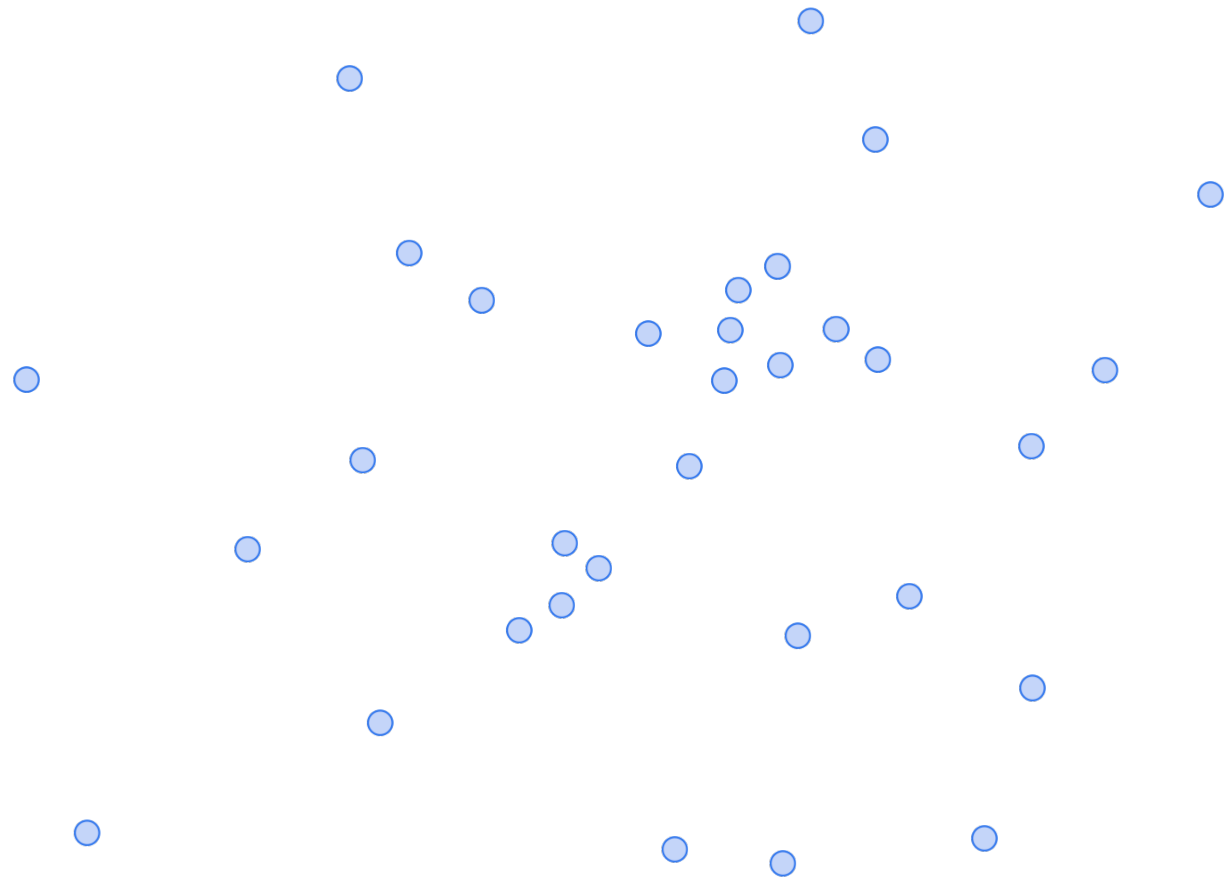
- The number of neighbouring cells that must be searched grows exponentially in the dimensionality in the worst case.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



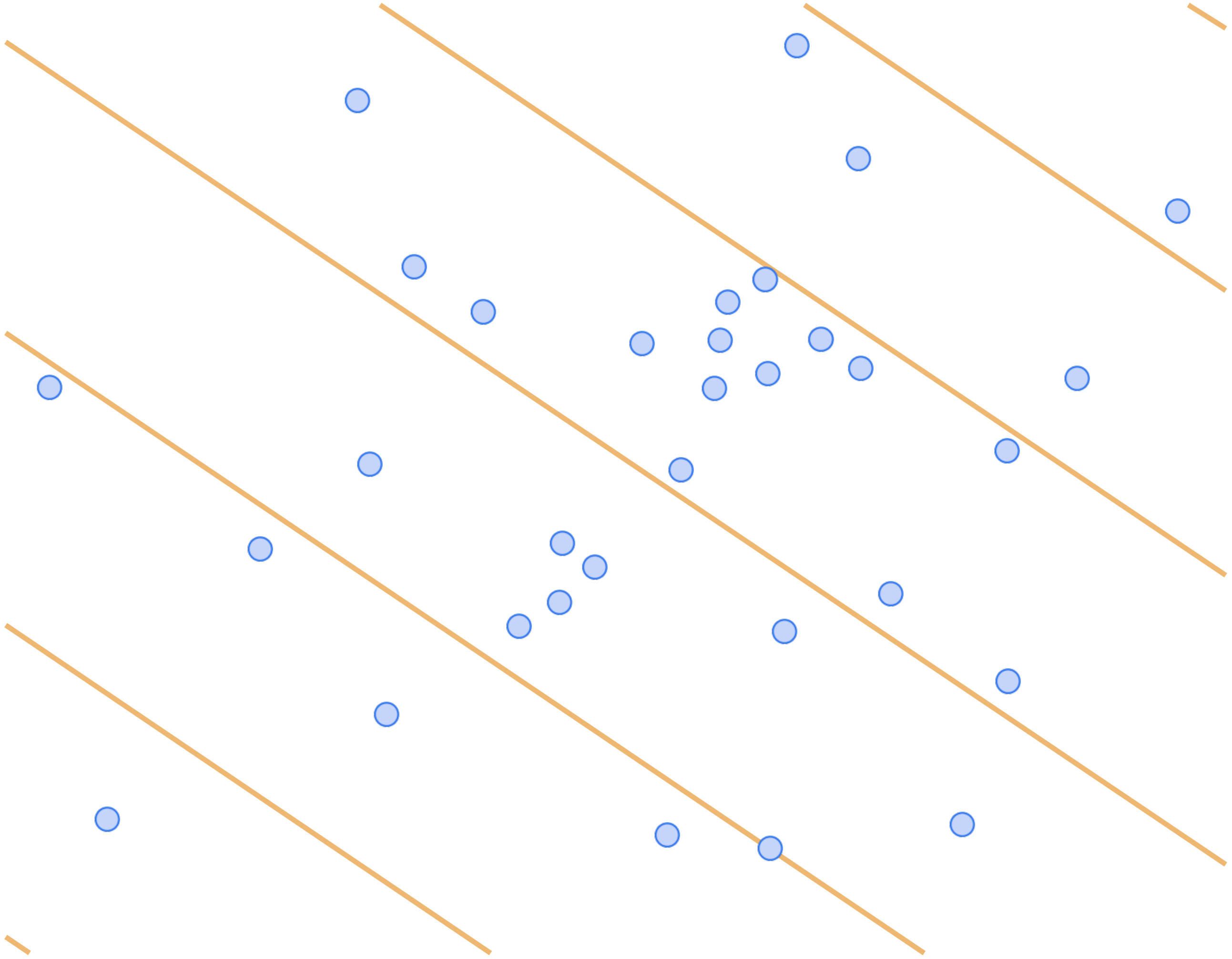
Berkeley
UNIVERSITY OF CALIFORNIA

# LSH



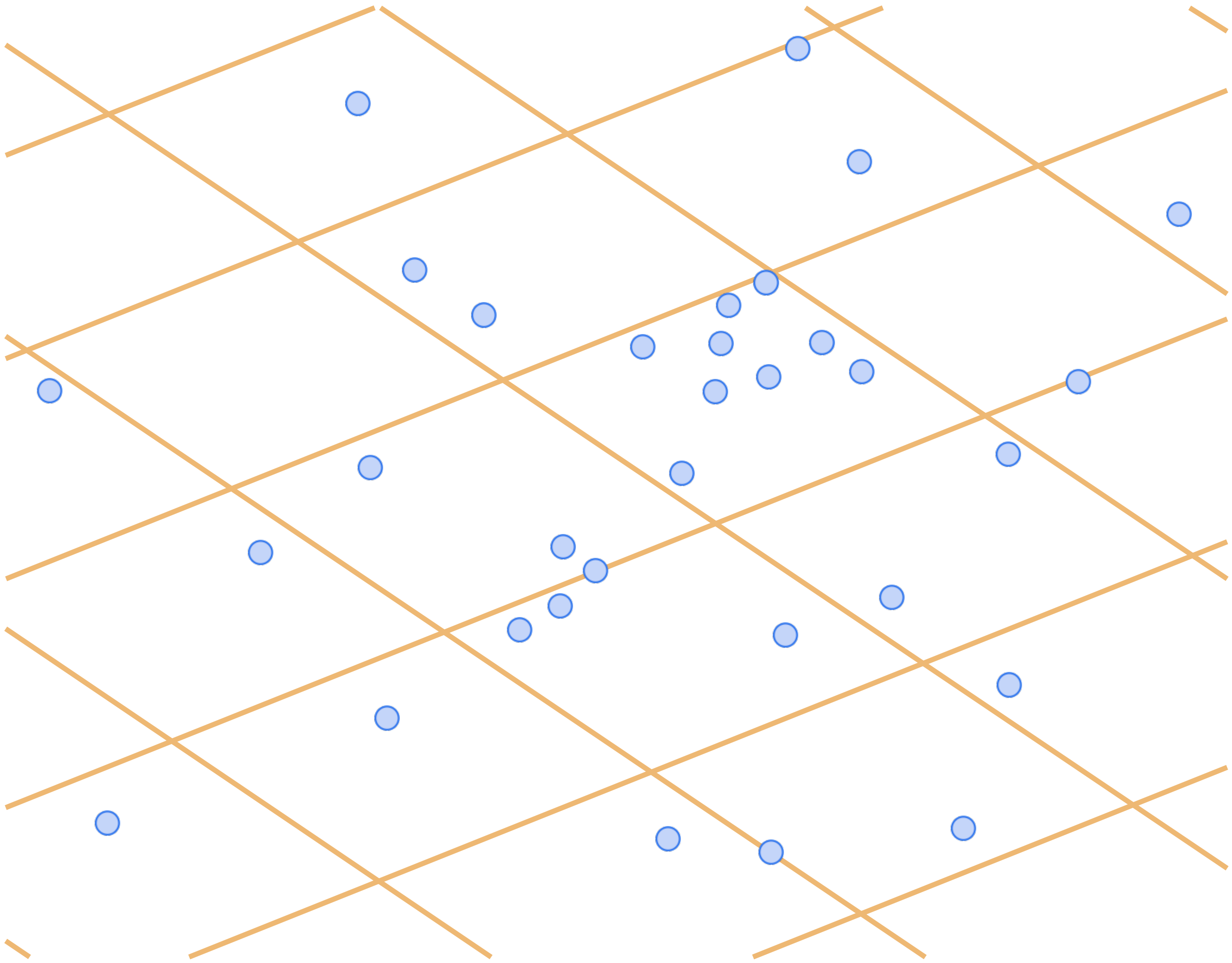
K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# LSH



K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# LSH

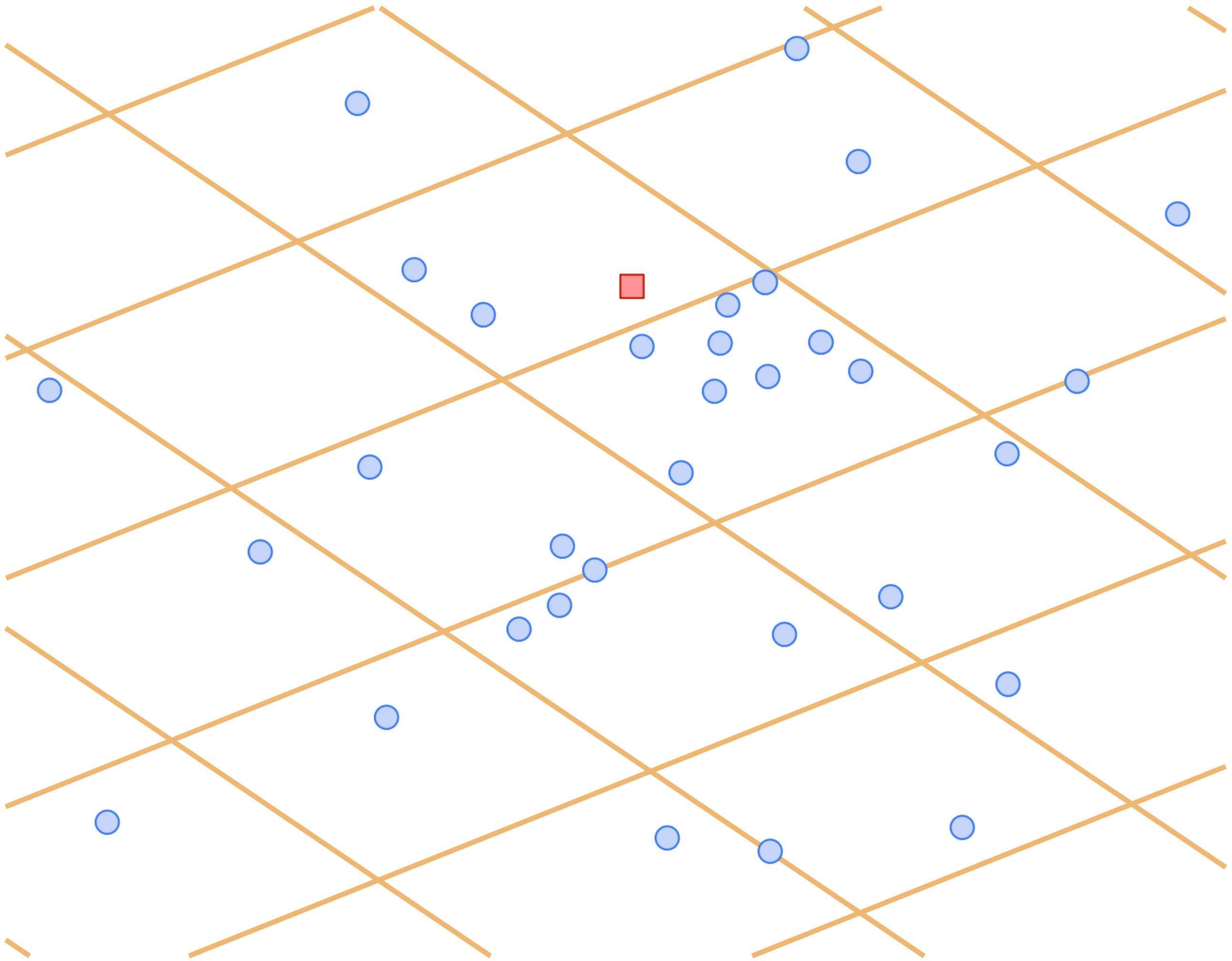


K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



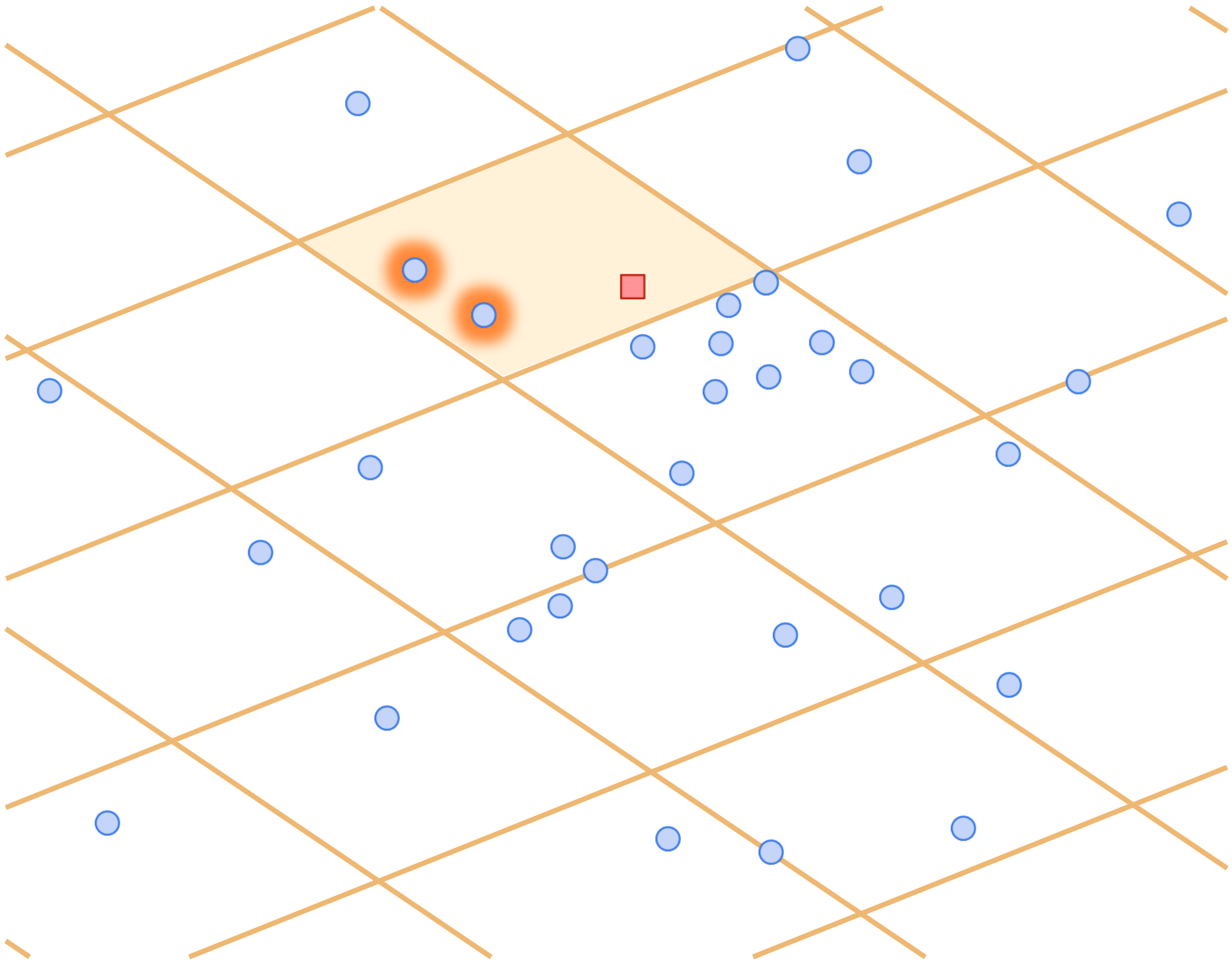
Berkeley
UNIVERSITY OF CALIFORNIA

# LSH



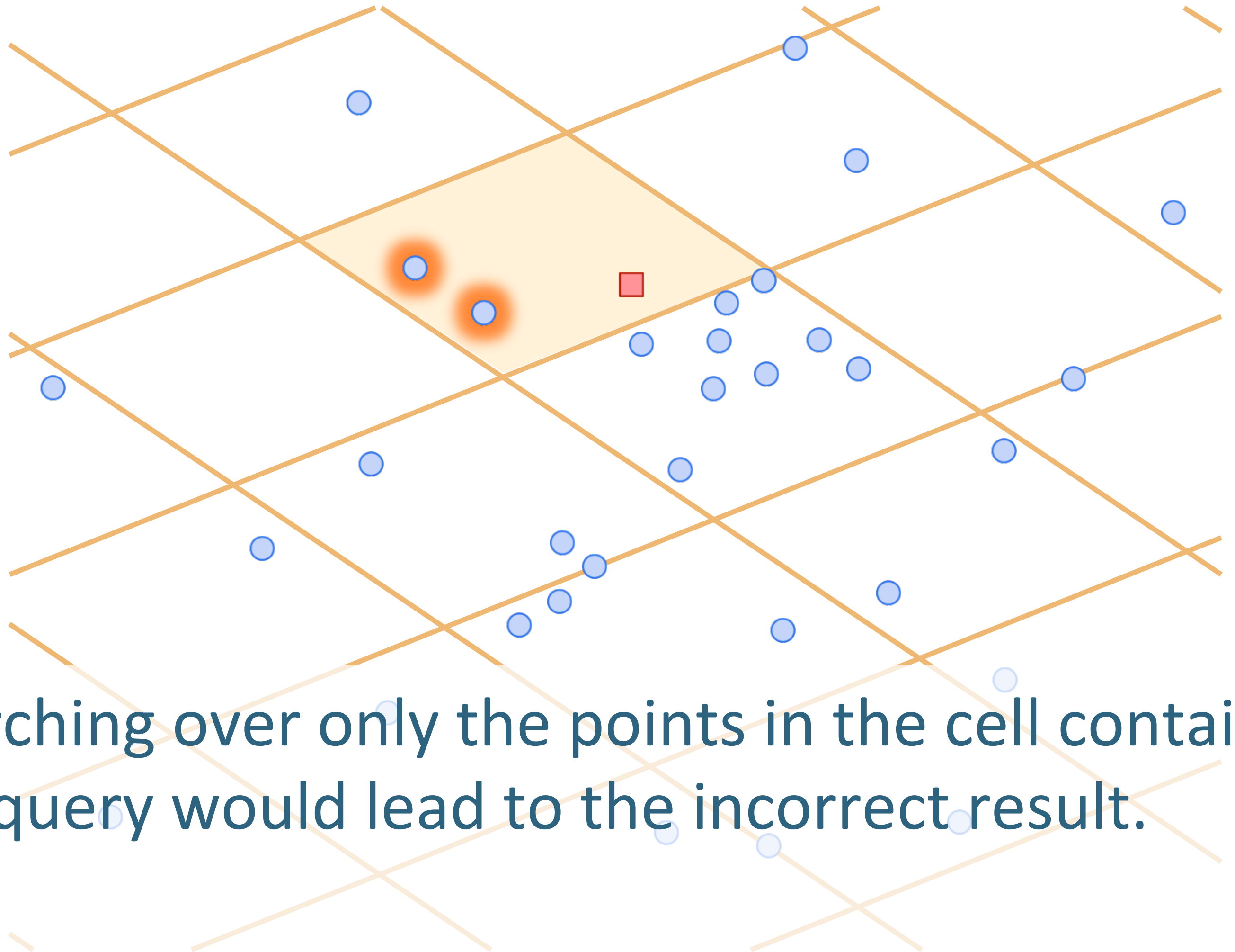
K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# LSH



K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

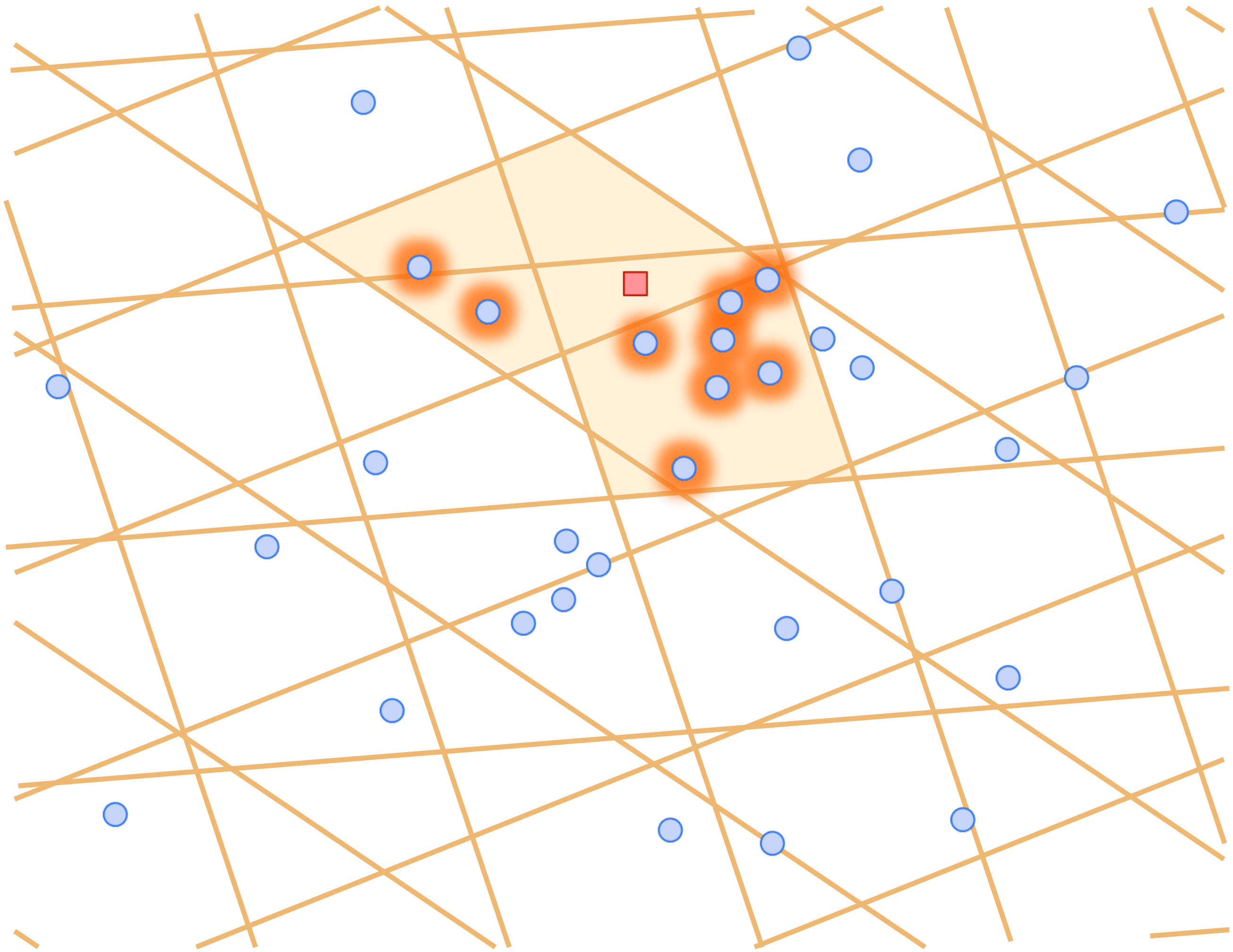
# LSH



- Searching over only the points in the cell containing the query would lead to the incorrect result.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# LSH

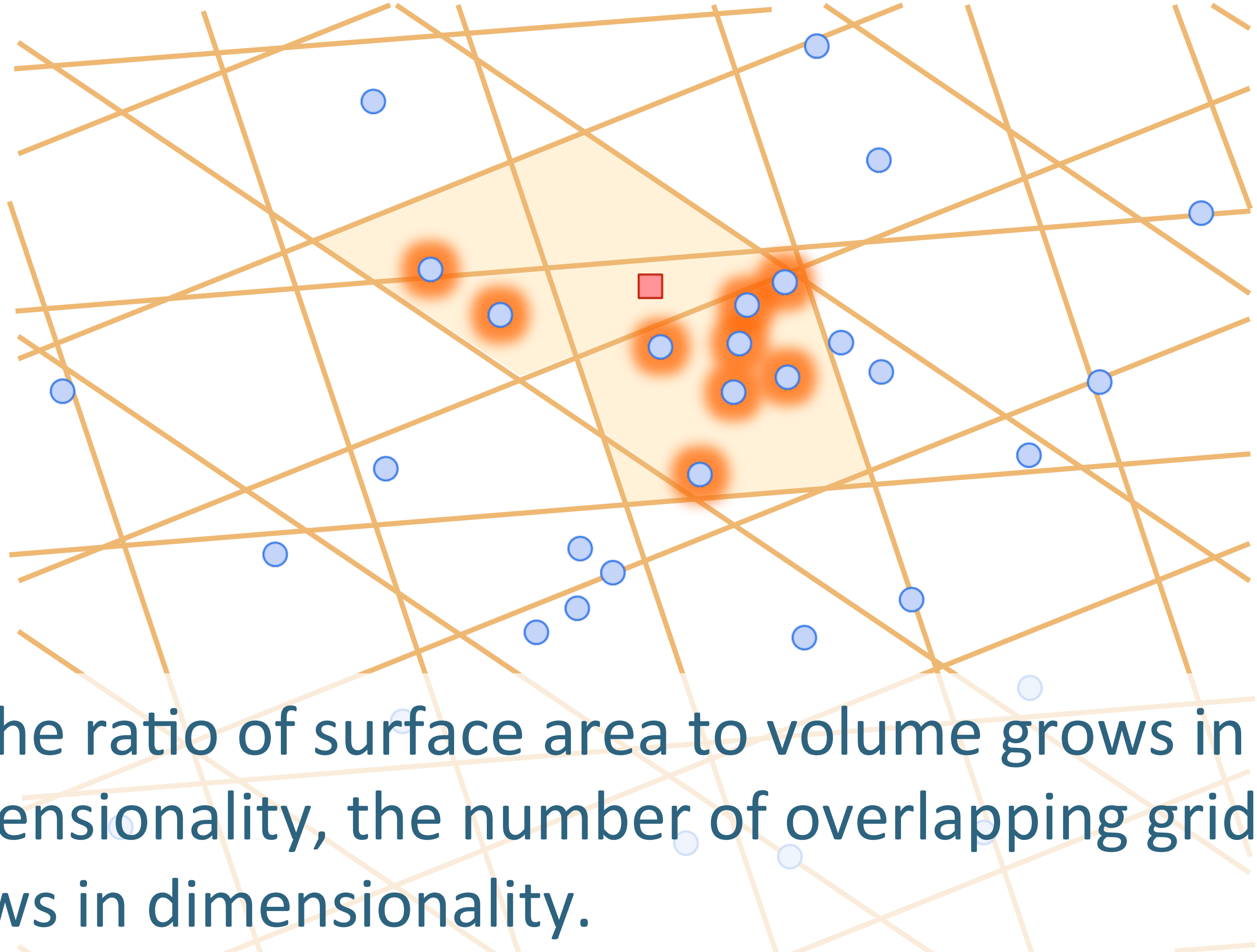


K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# LSH



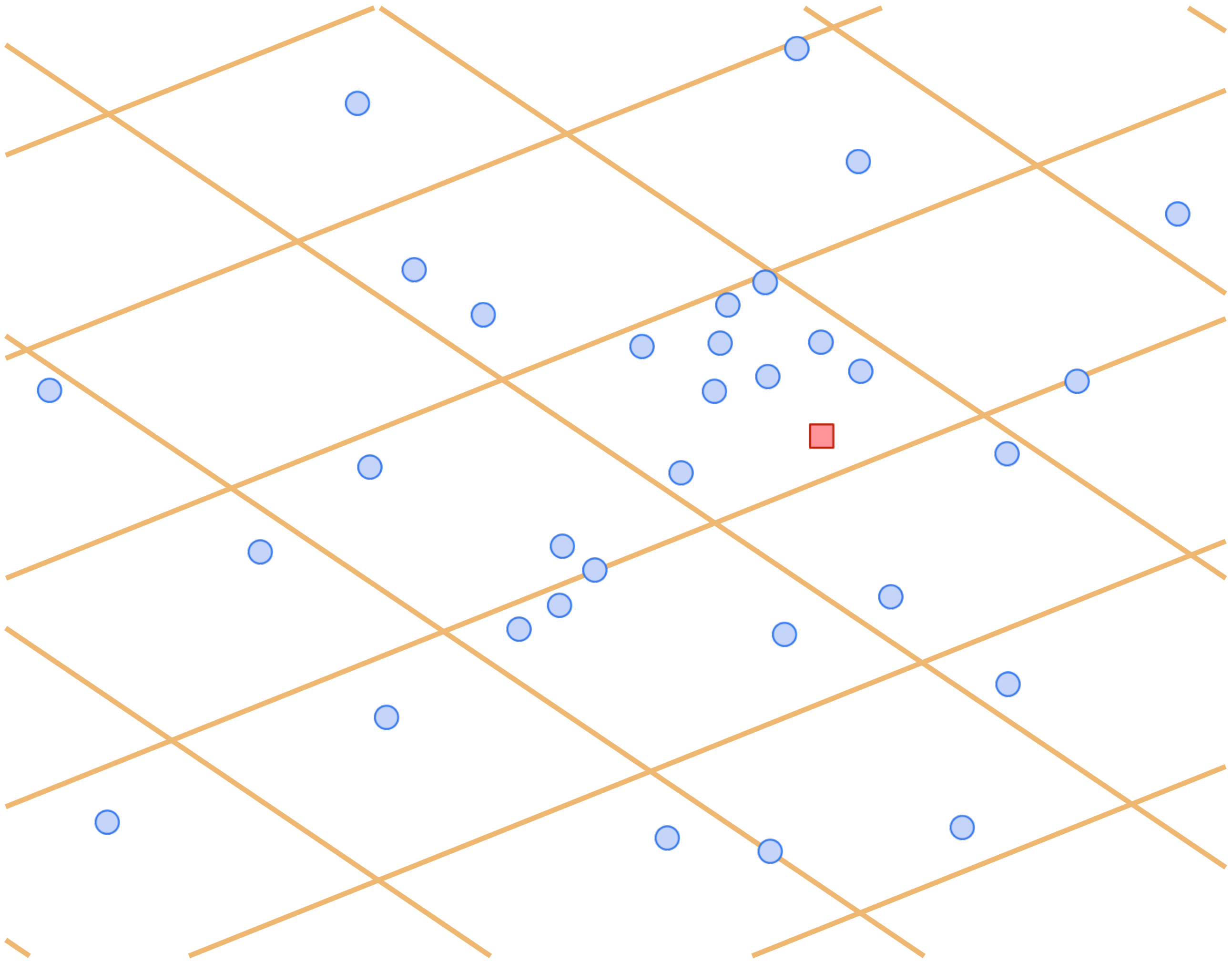
- As the ratio of surface area to volume grows in dimensionality, the number of overlapping grids grows in dimensionality.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



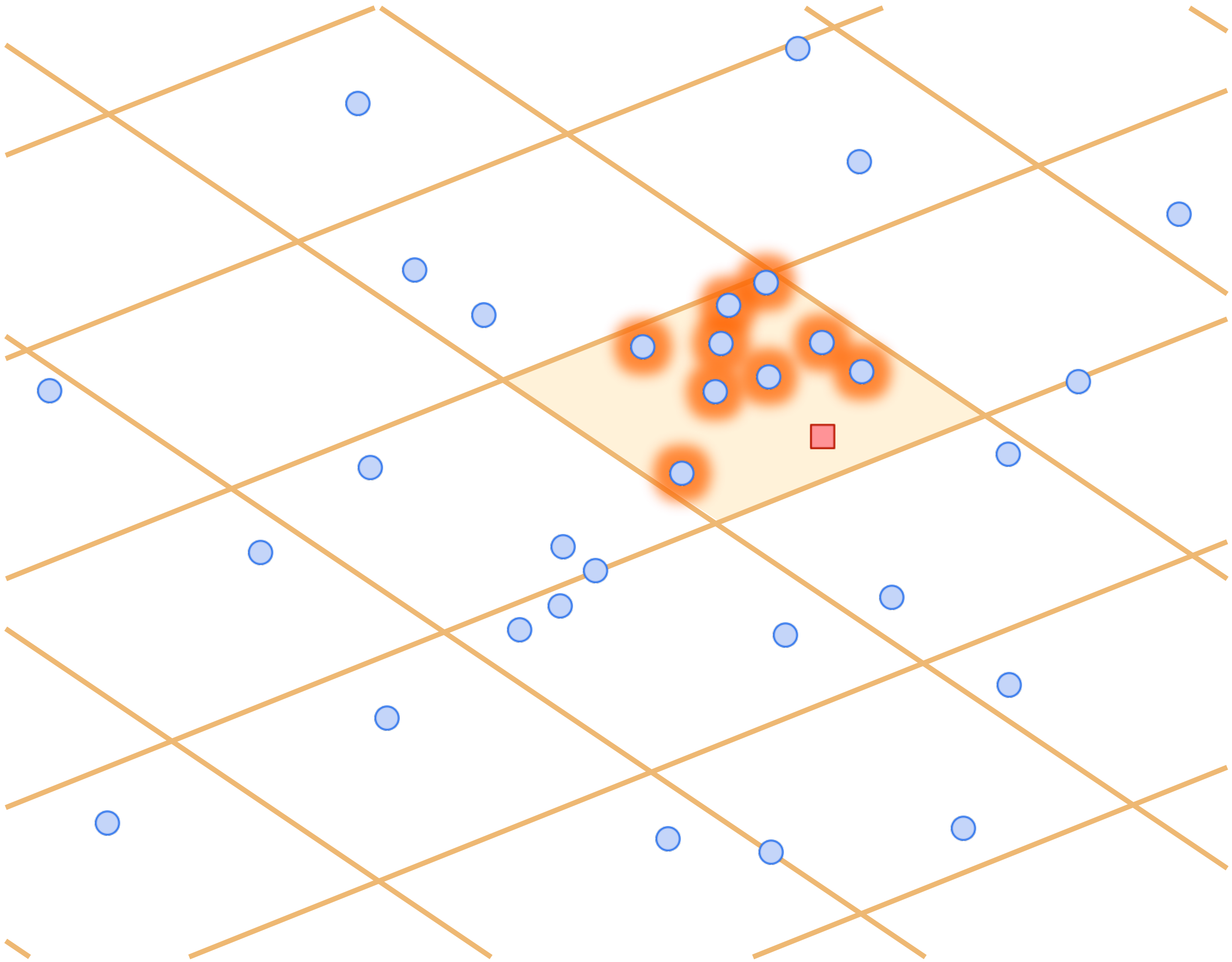
Berkeley
UNIVERSITY OF CALIFORNIA

# LSH



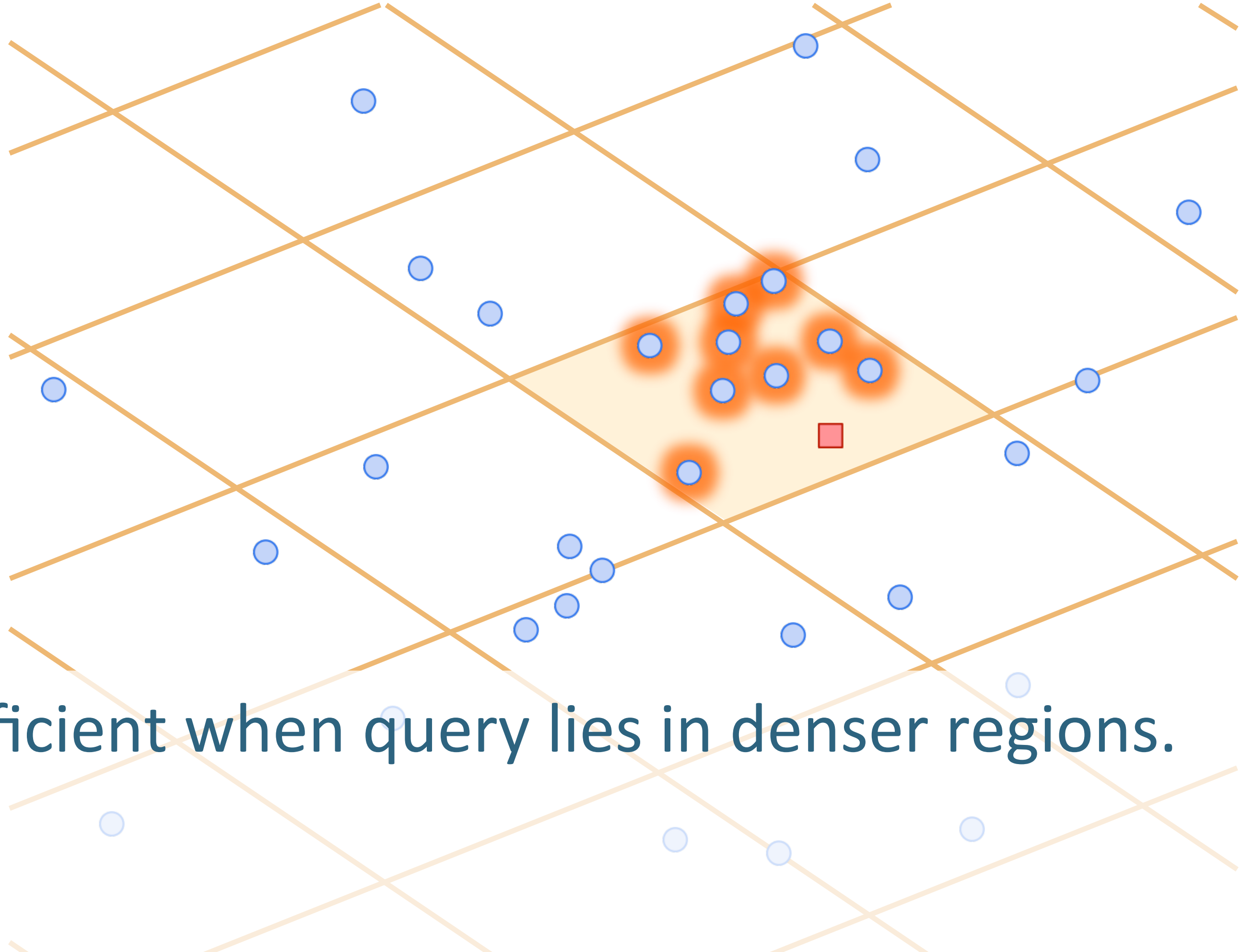
K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# LSH



K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# LSH



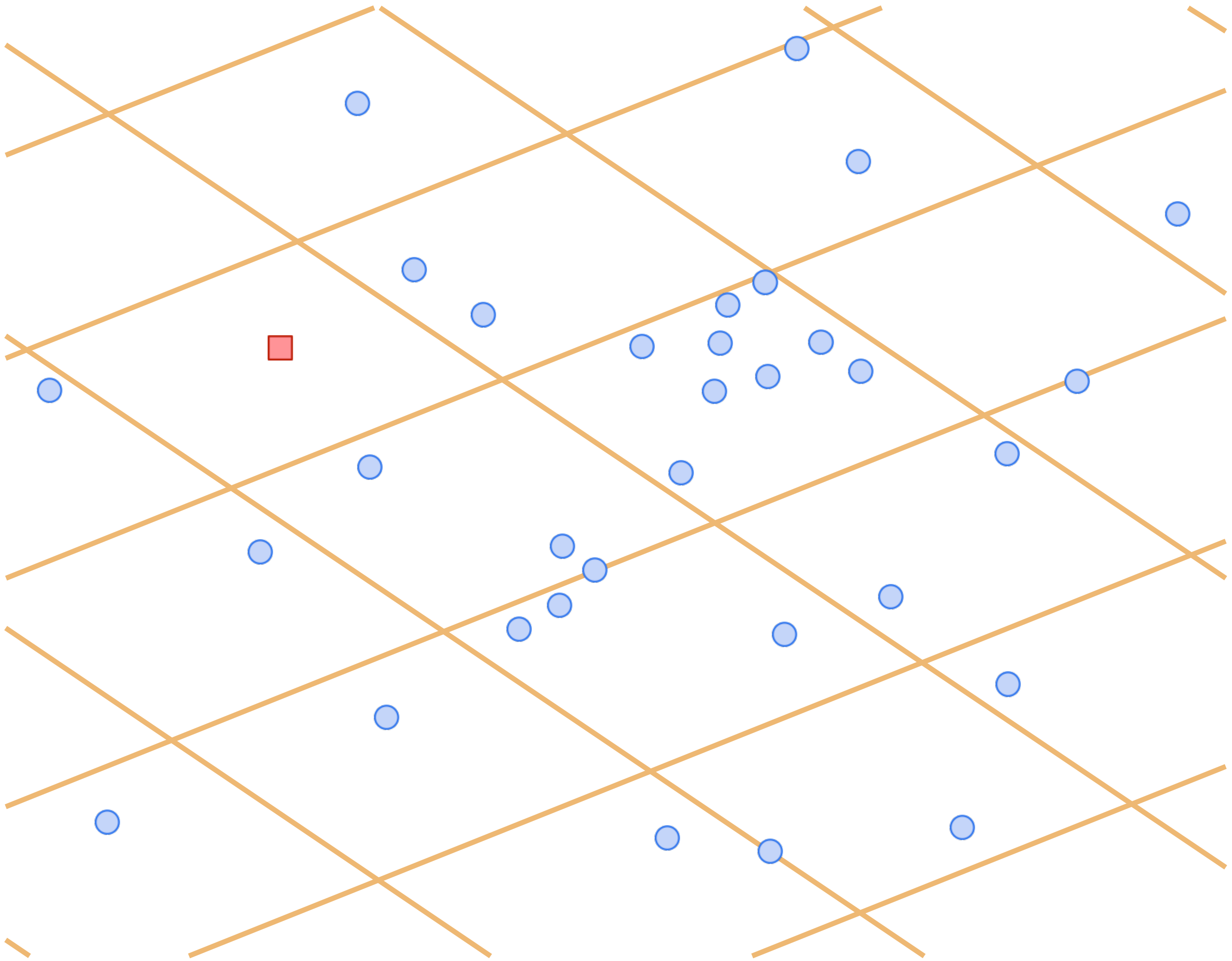
- Inefficient when query lies in denser regions.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



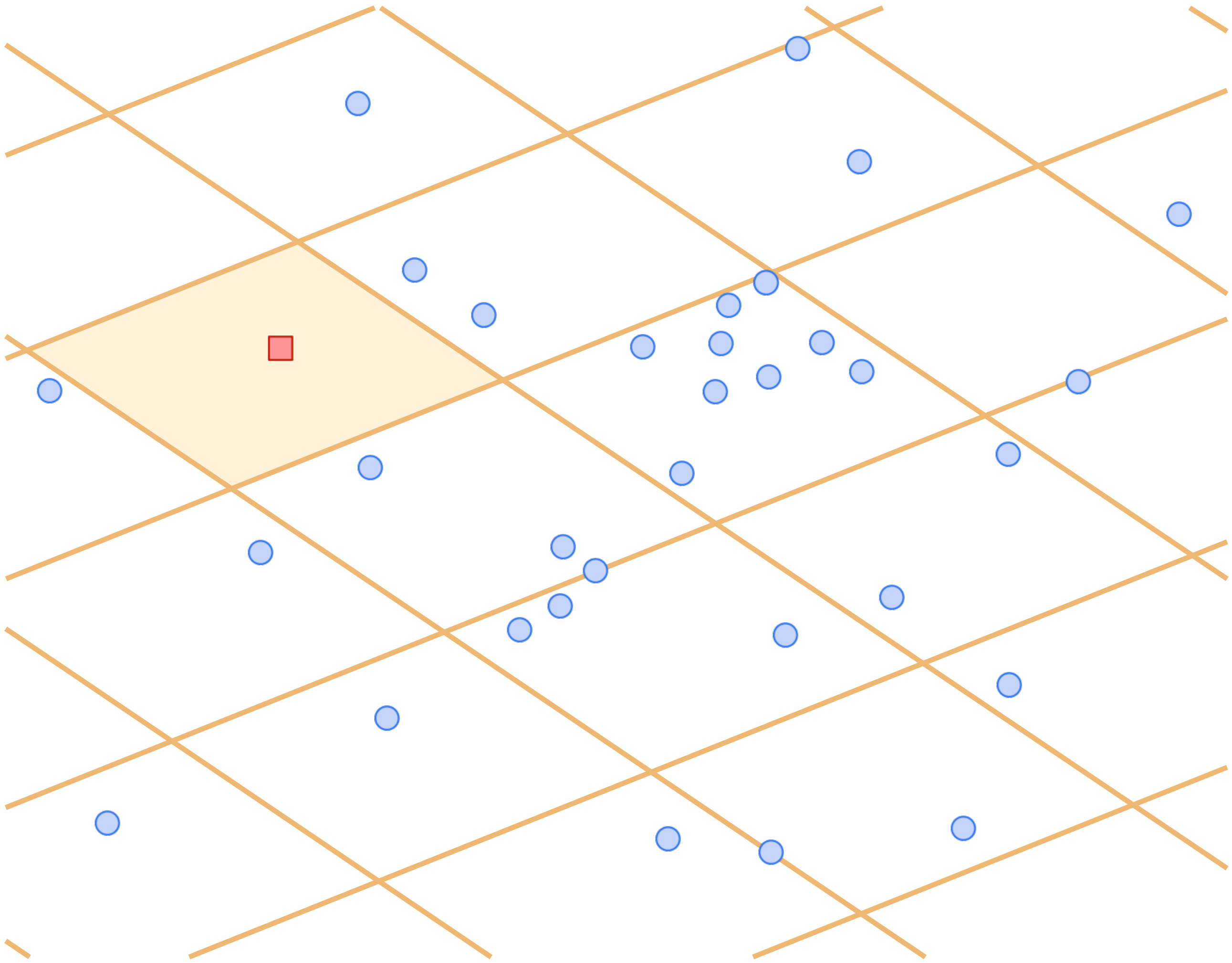
Berkeley
UNIVERSITY OF CALIFORNIA

# LSH



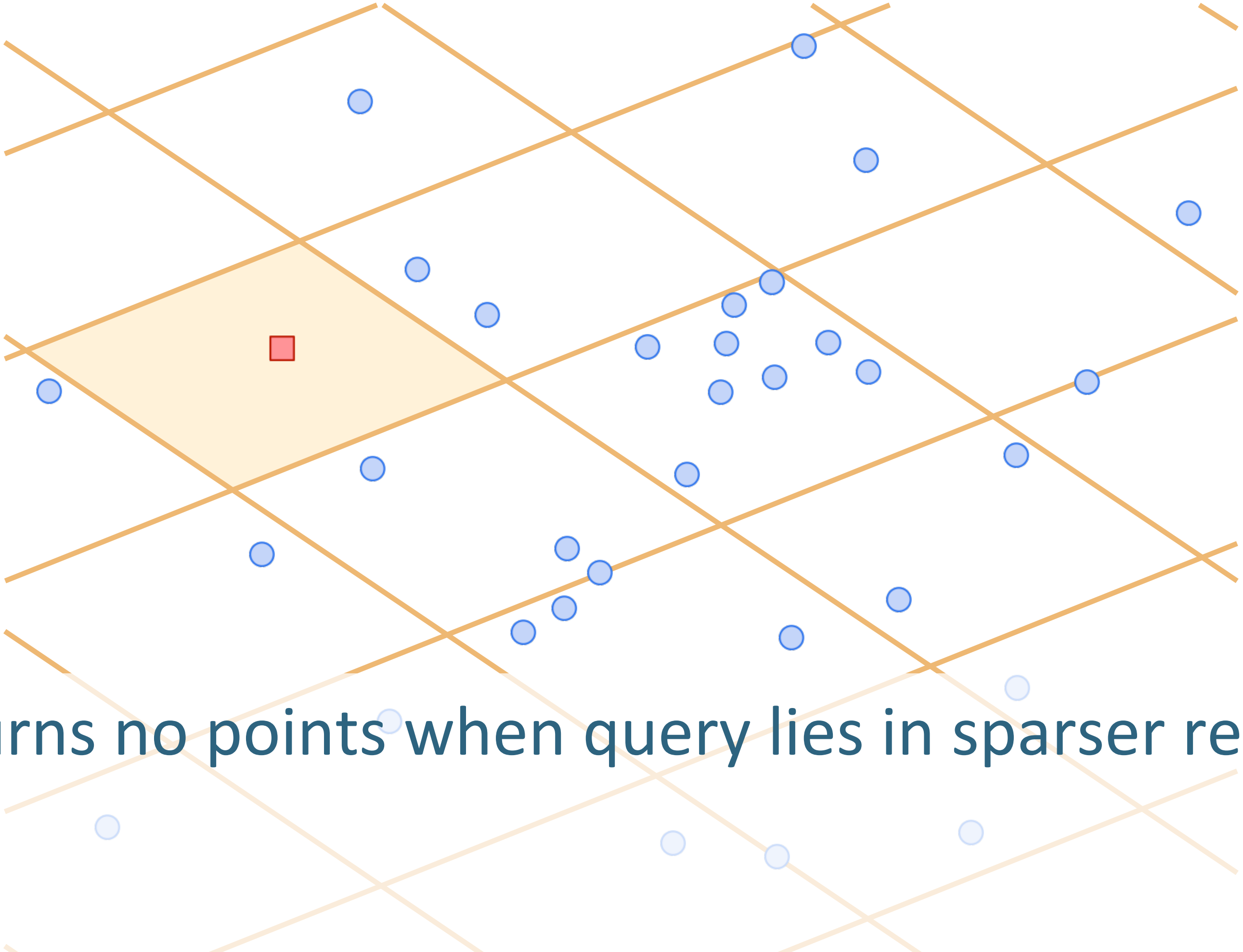
K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# LSH



K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# LSH



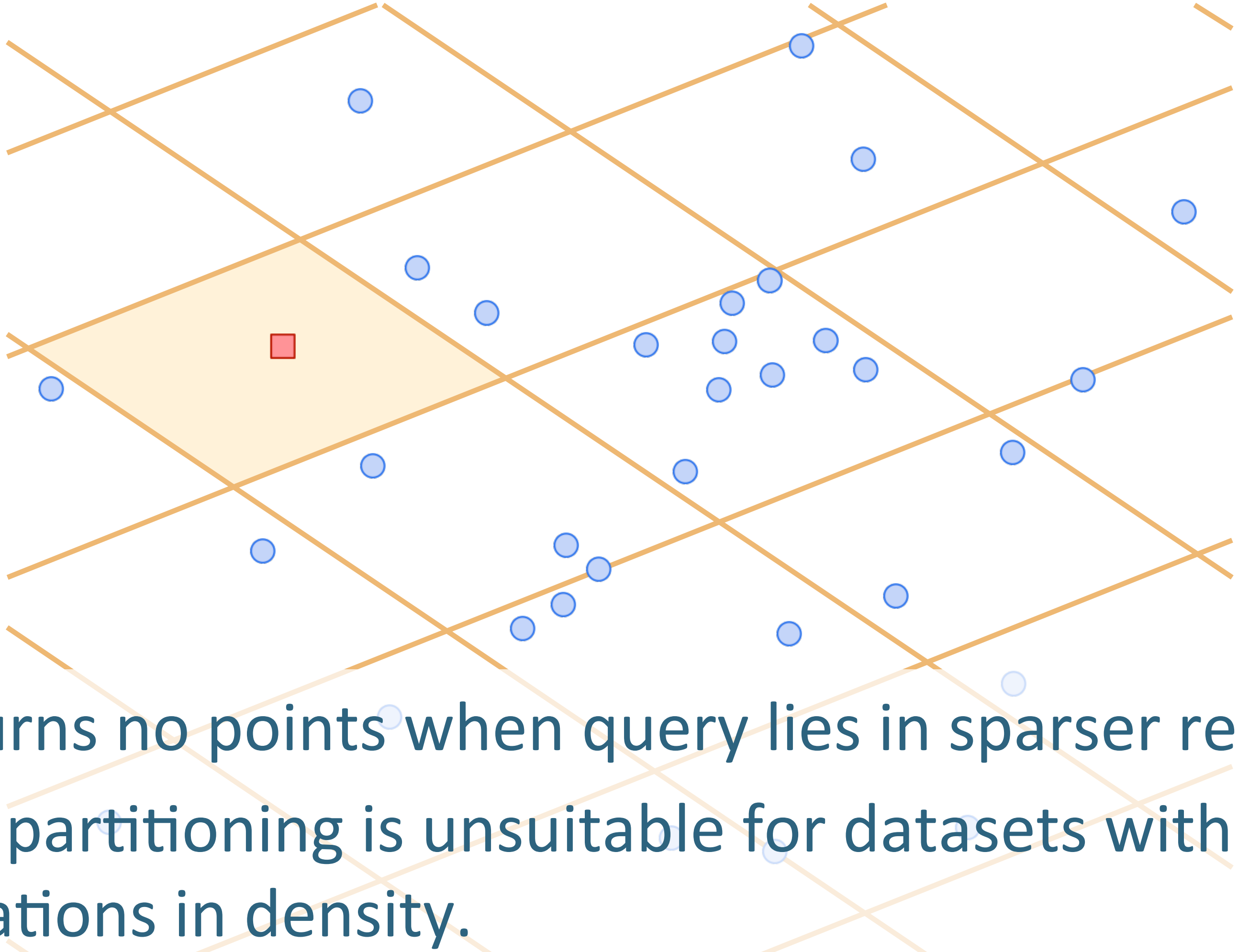
- Returns no points when query lies in sparser regions.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# LSH



- Returns no points when query lies in sparser regions.
- This partitioning is unsuitable for datasets with large variations in density.

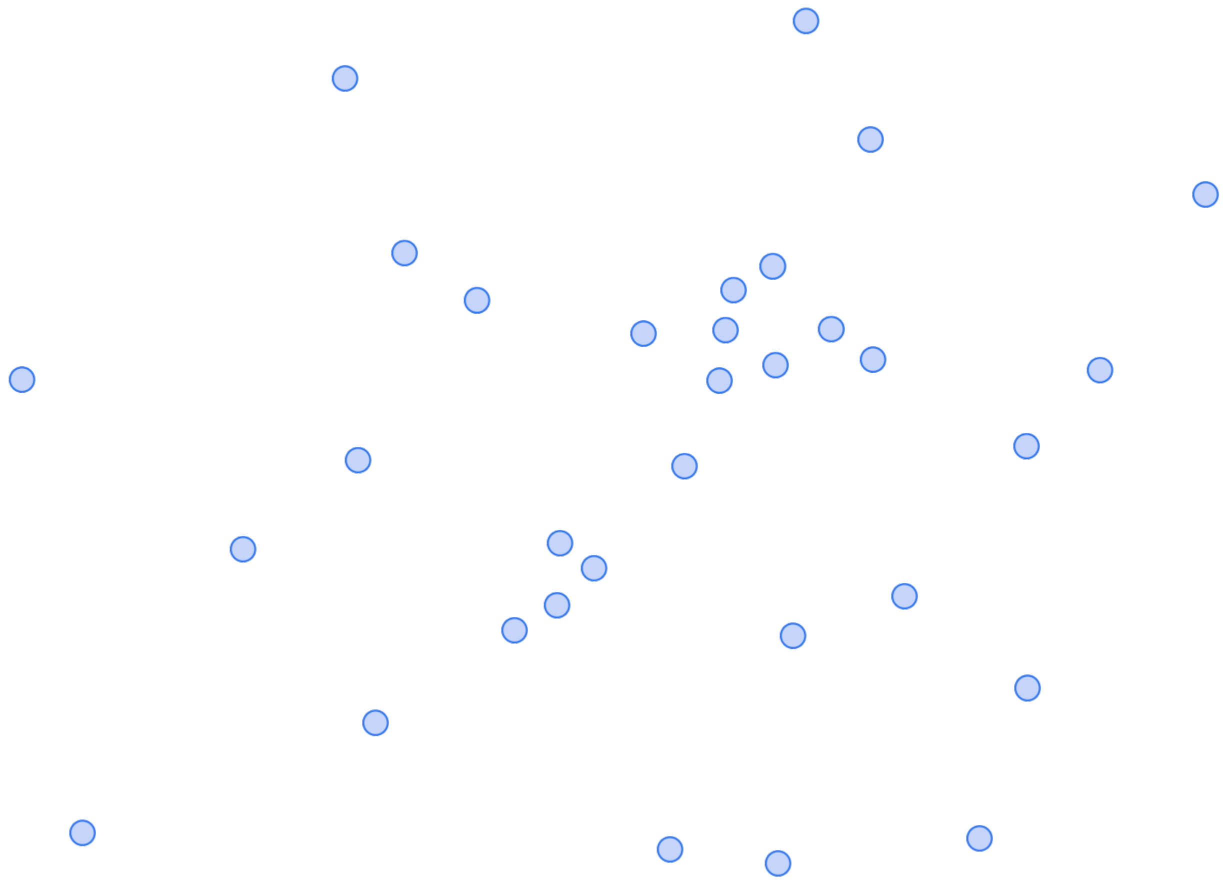
K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

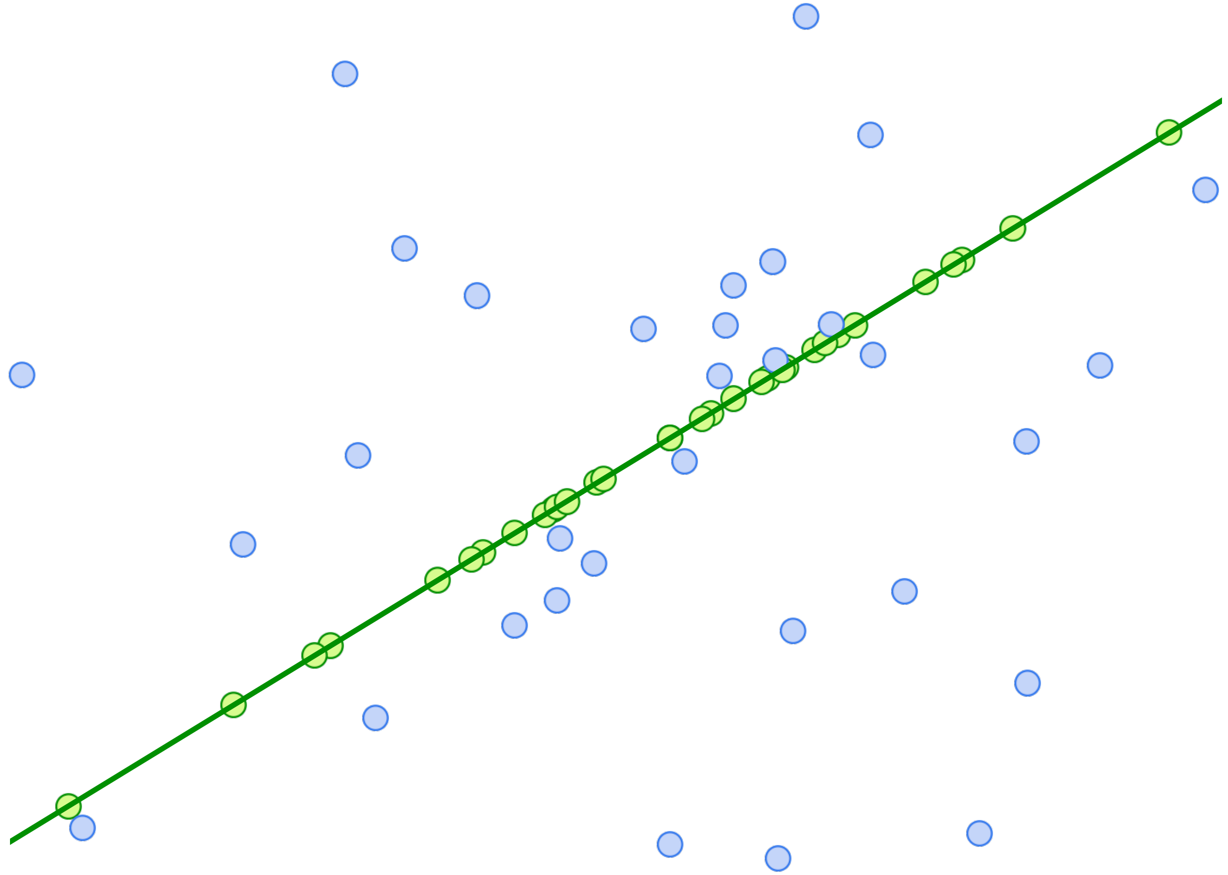
# Prioritized DCI

# Prioritized DCI



K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

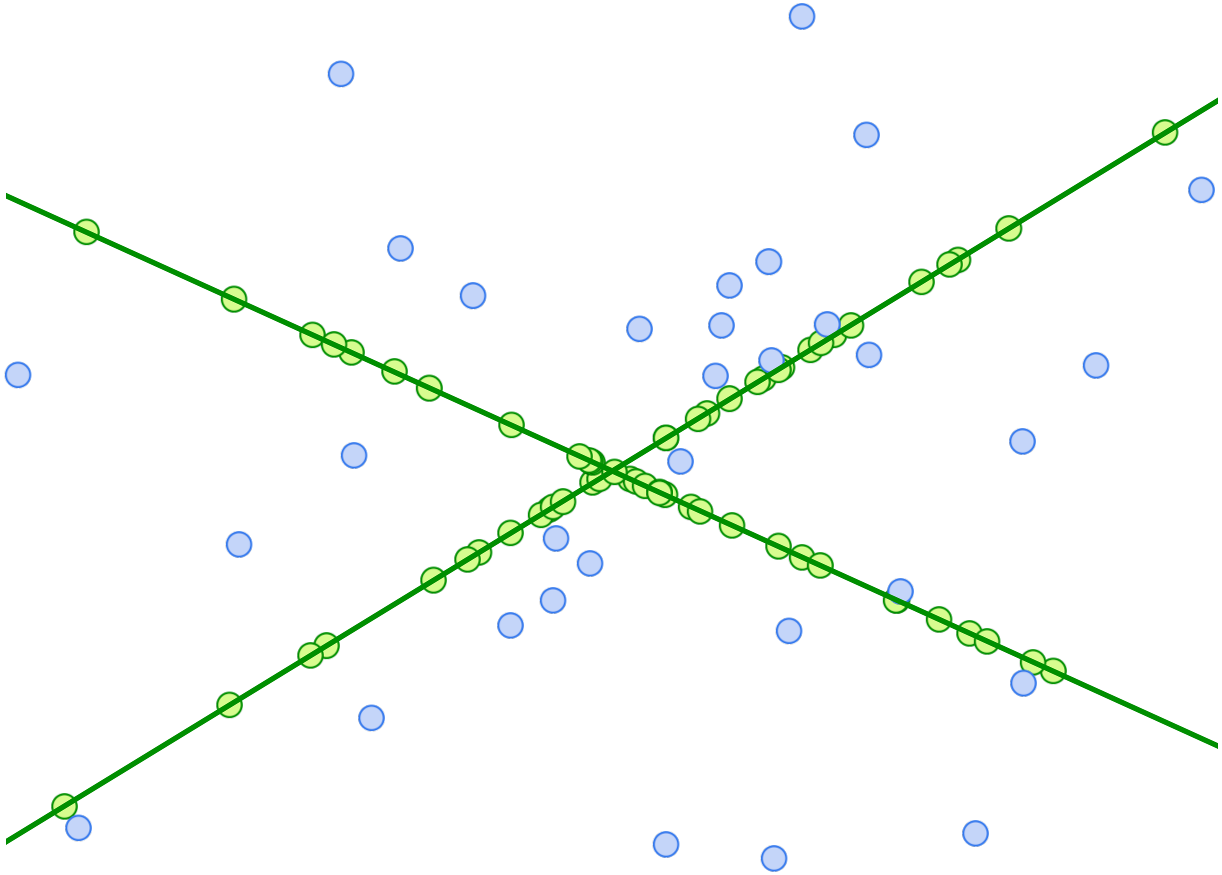
# Prioritized DCI



Project all data points along a random direction.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# Prioritized DCI



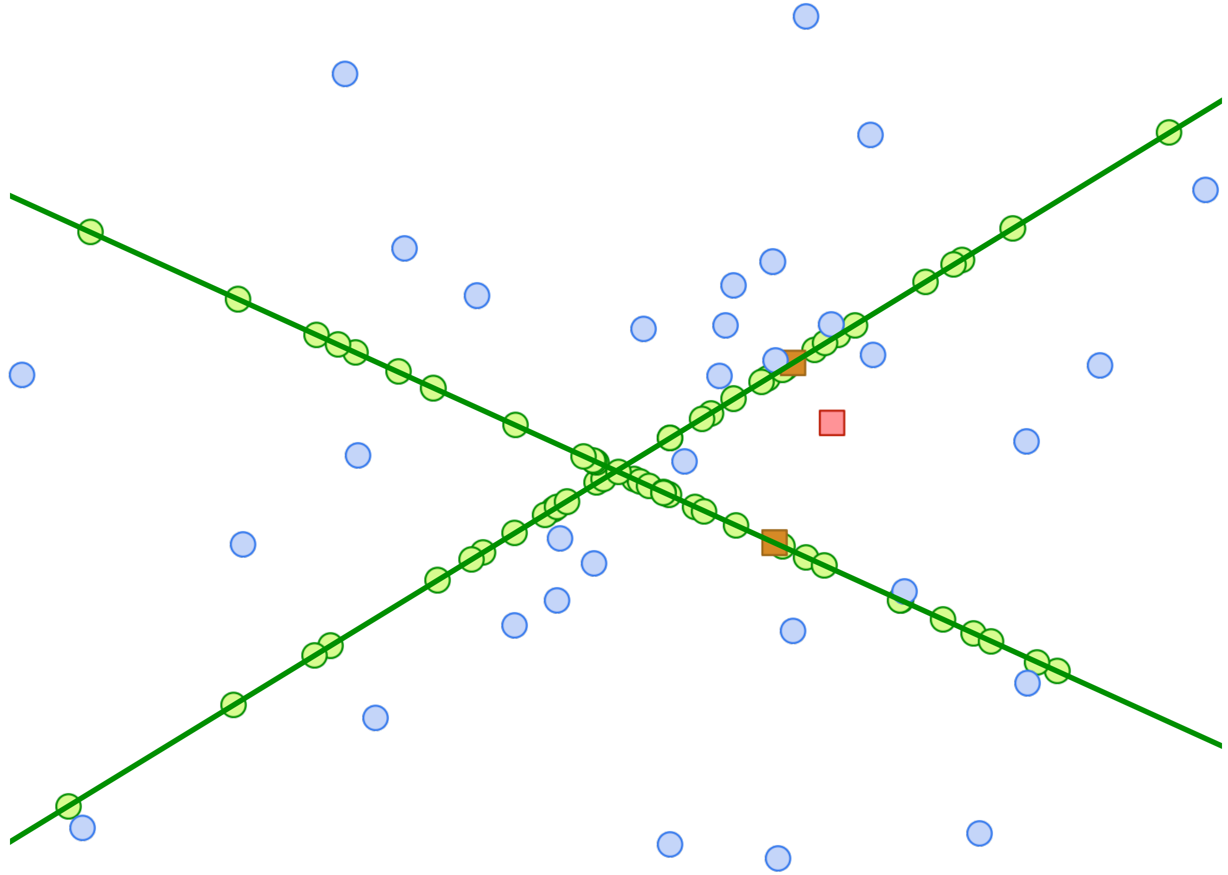
Project all data points along multiple random directions.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI



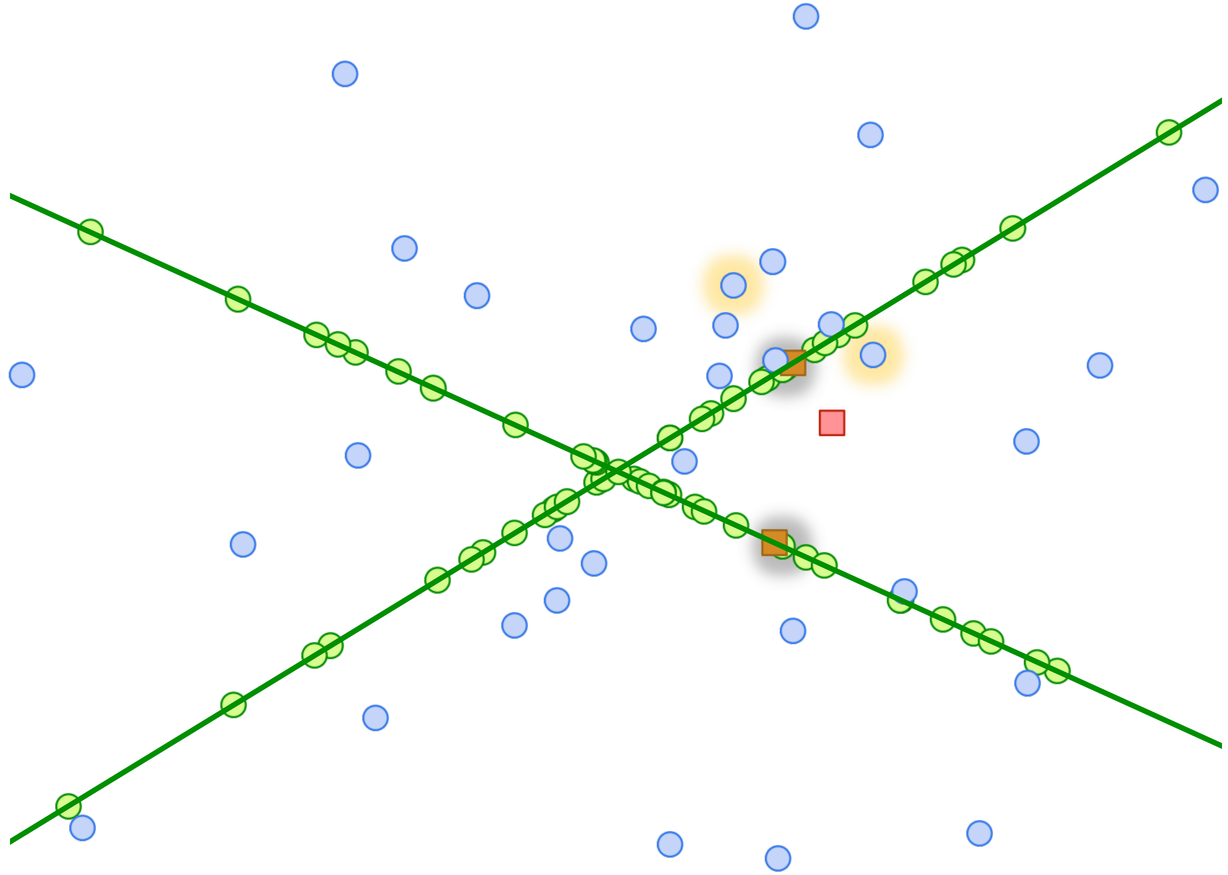
Project the query along each projection direction.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI



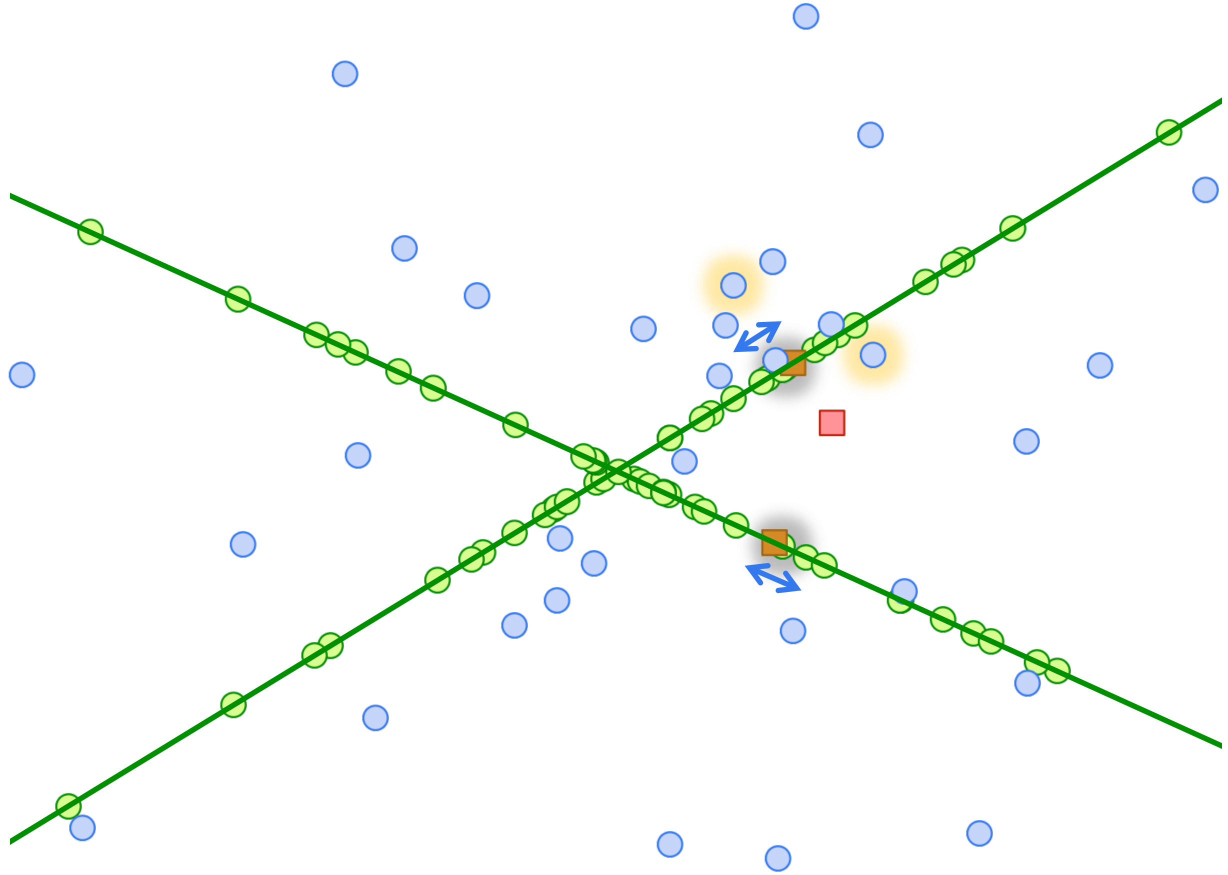
Find the closest point to the query along each projection direction and add them to the frontier.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

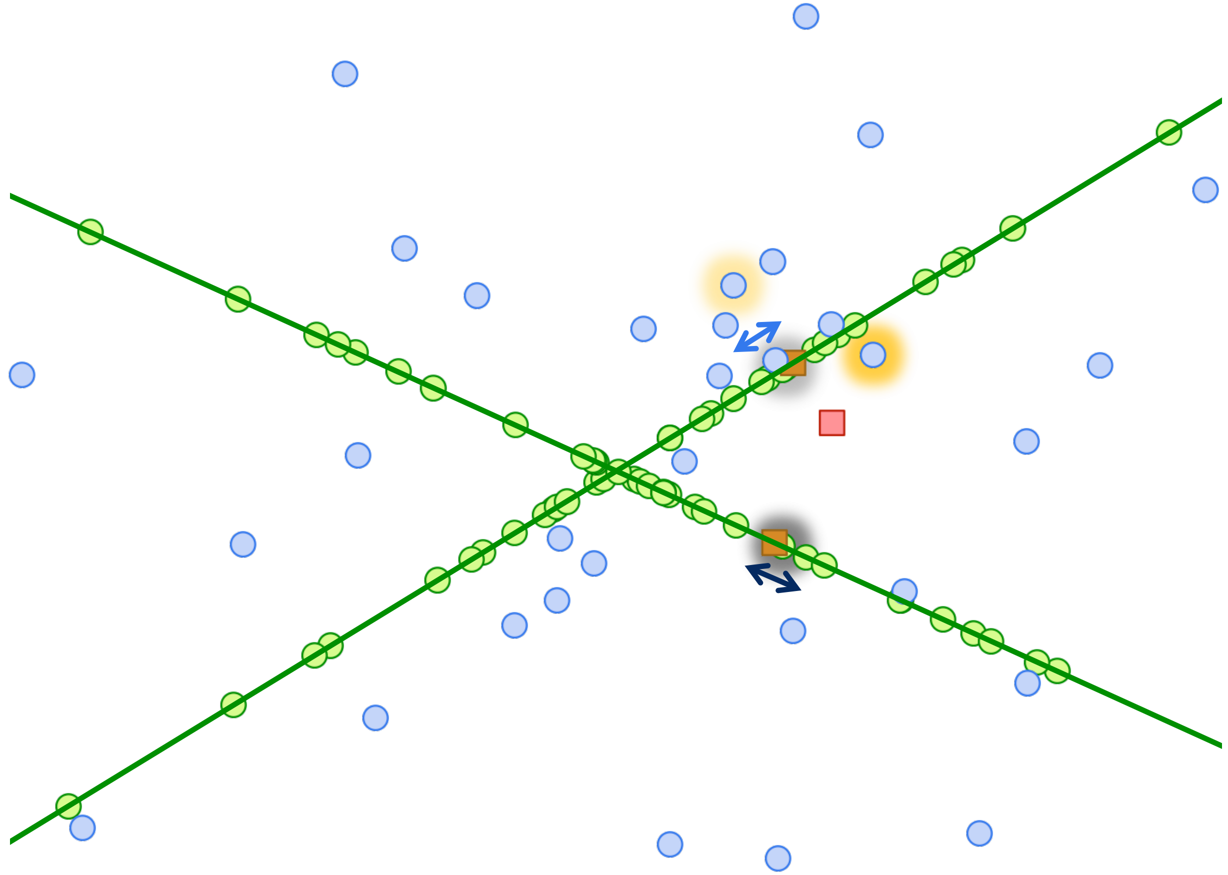
# Prioritized DCI



Compare their projected distances to the query.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# Prioritized DCI



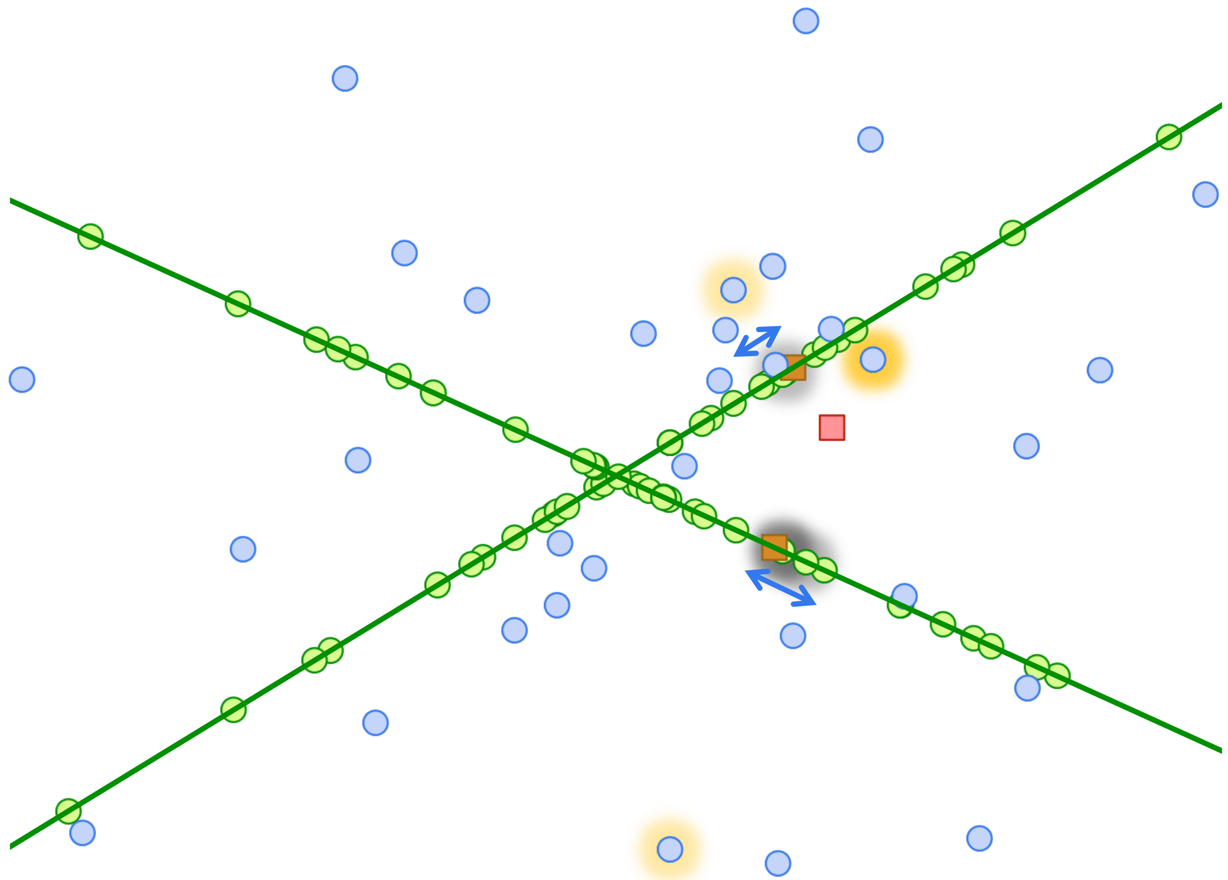
Visit the point with the shortest projected distance.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI

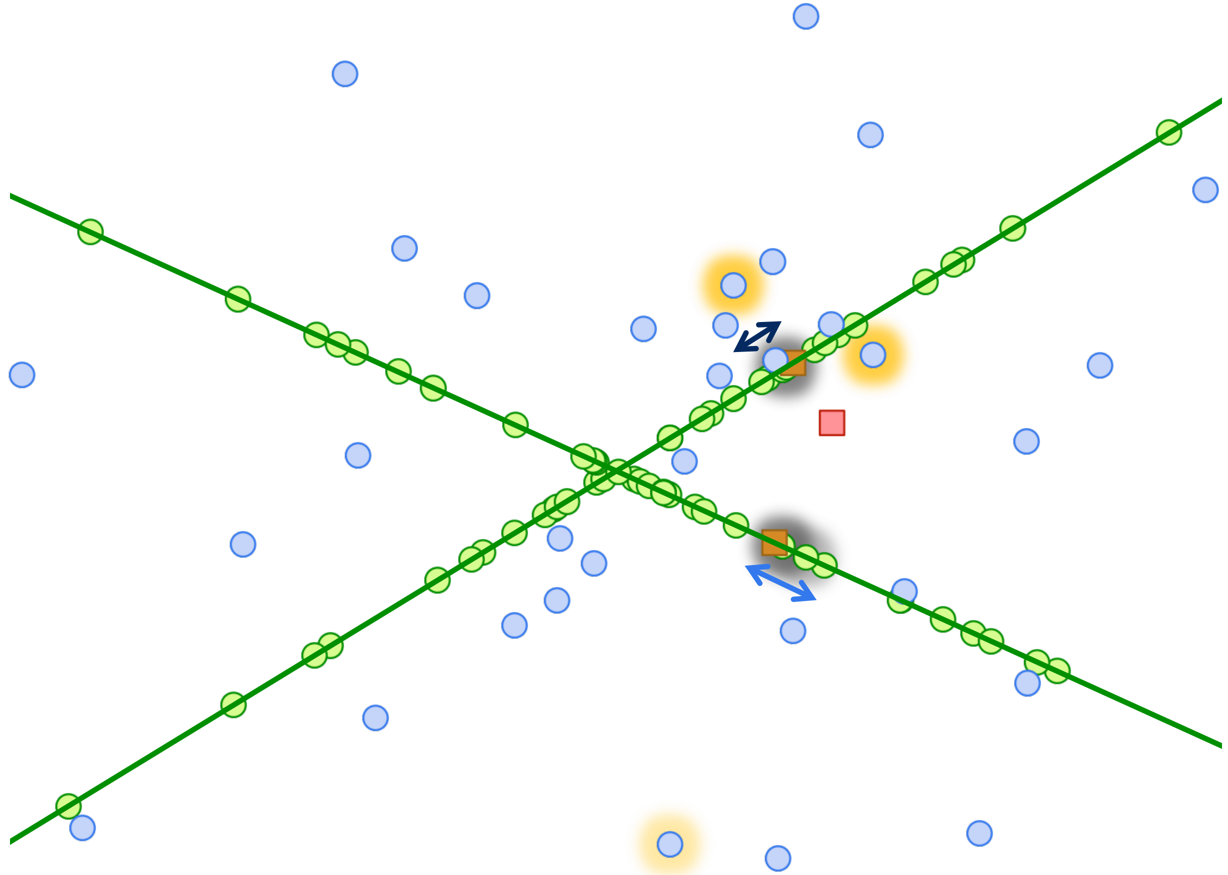


Find the next closest point along the projection direction that has just been processed and add it to the frontier.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI



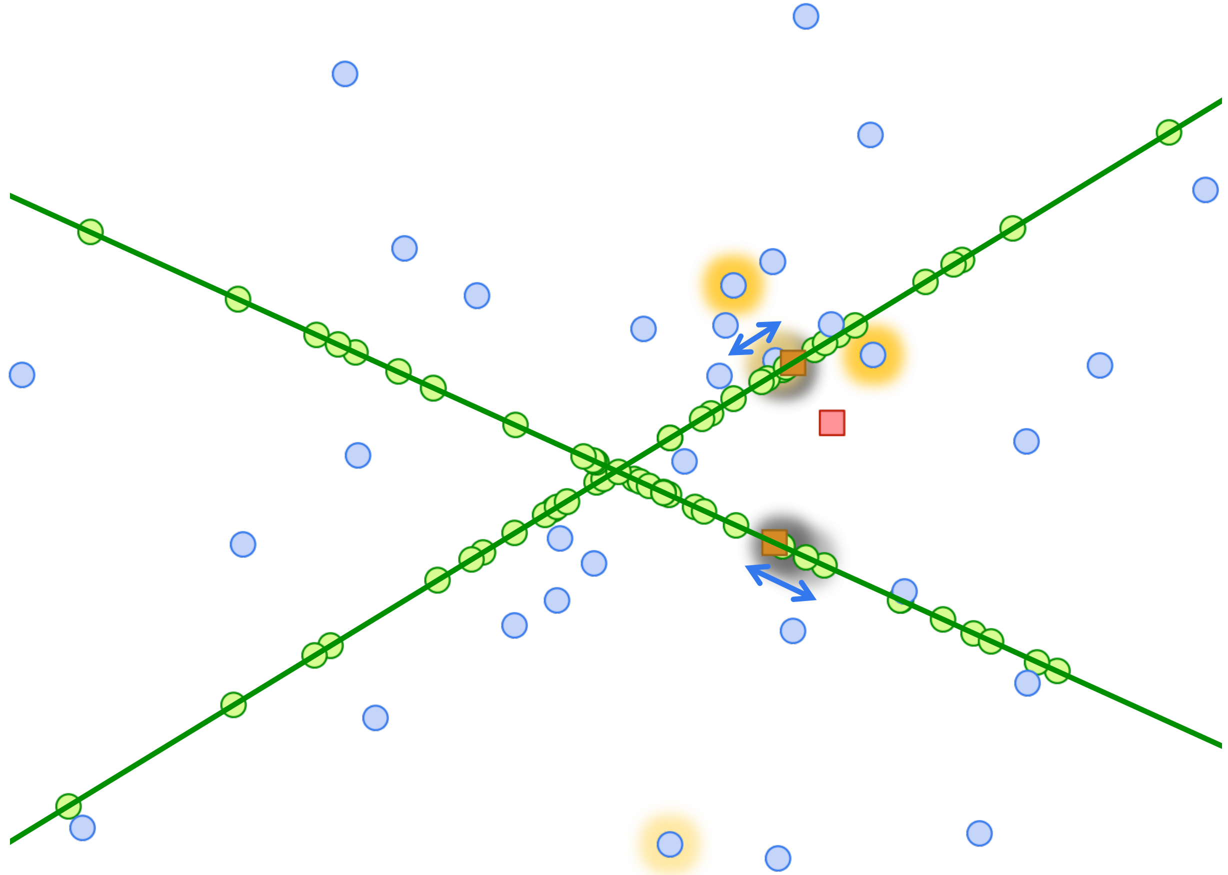
Compare projected distances of points on the frontier and visit the one with the shortest projected distance.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI



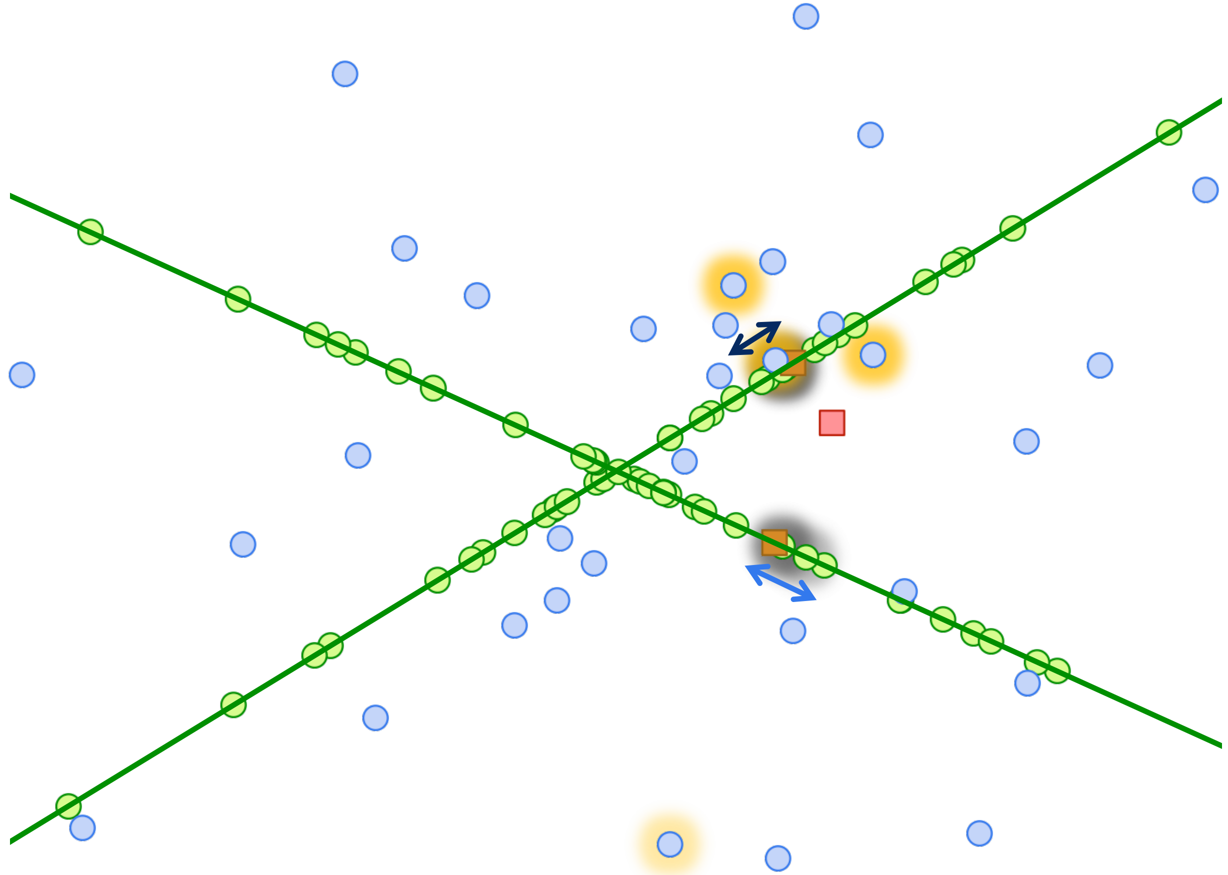
Find the next closest point along the projection direction that has just been processed and add it to the frontier.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI



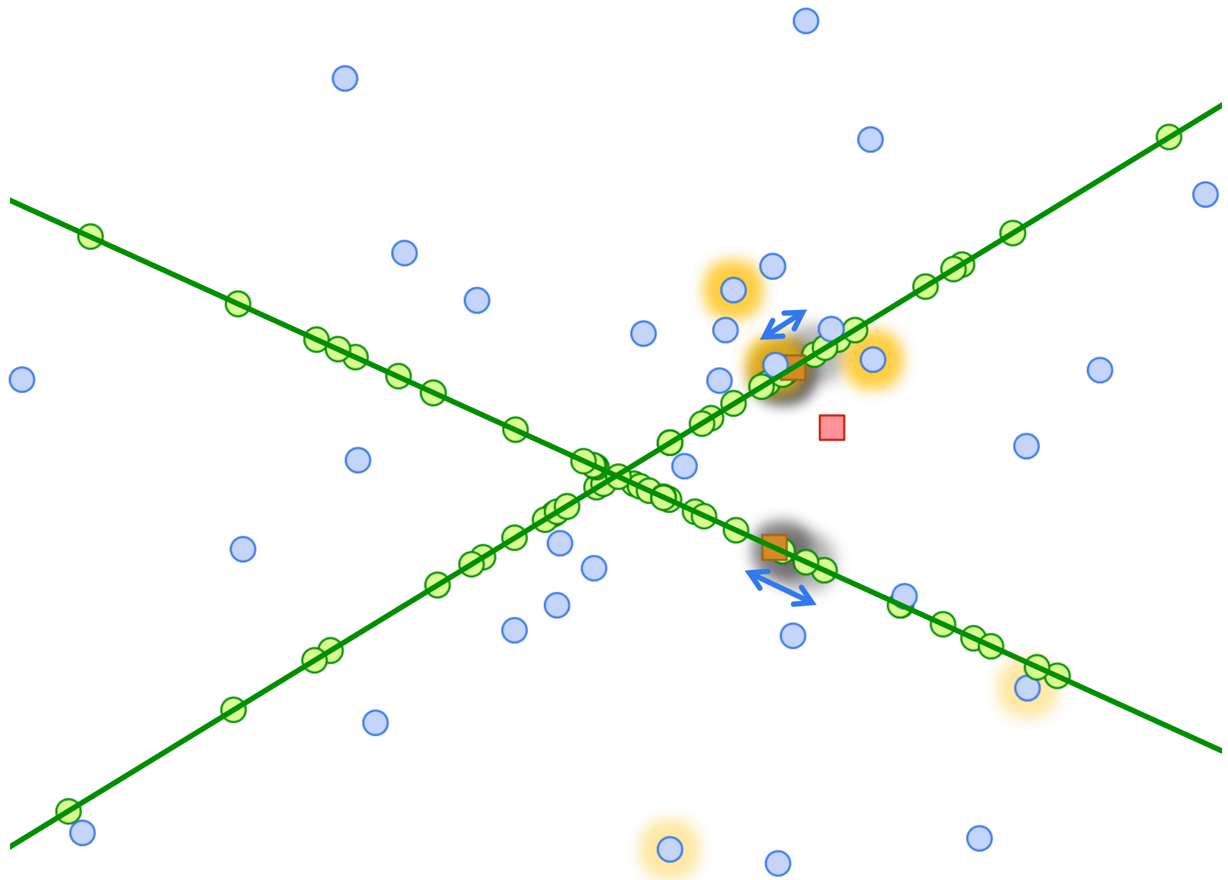
Compare projected distances of points on the frontier and visit the one with the shortest projected distance.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI



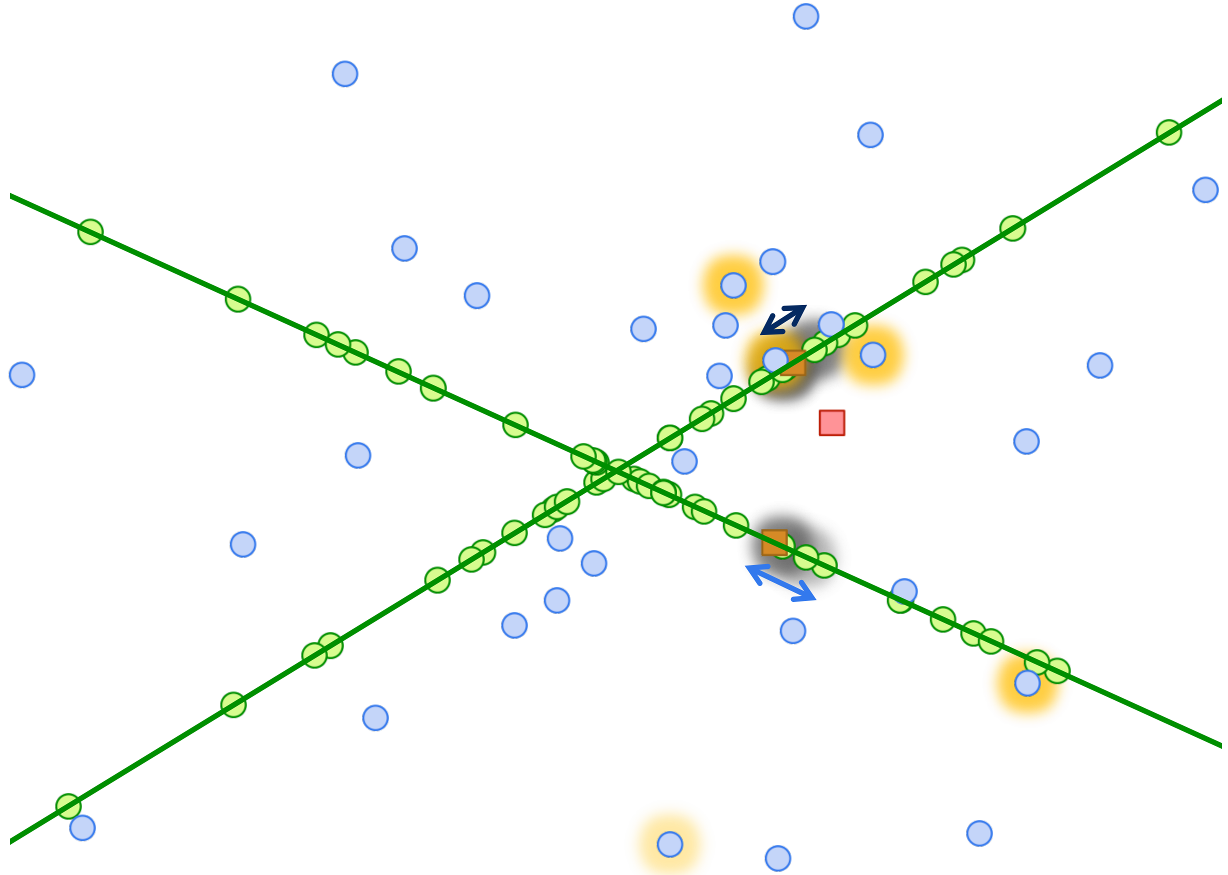
Find the next closest point along the projection direction that has just been processed and add it to the frontier.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI



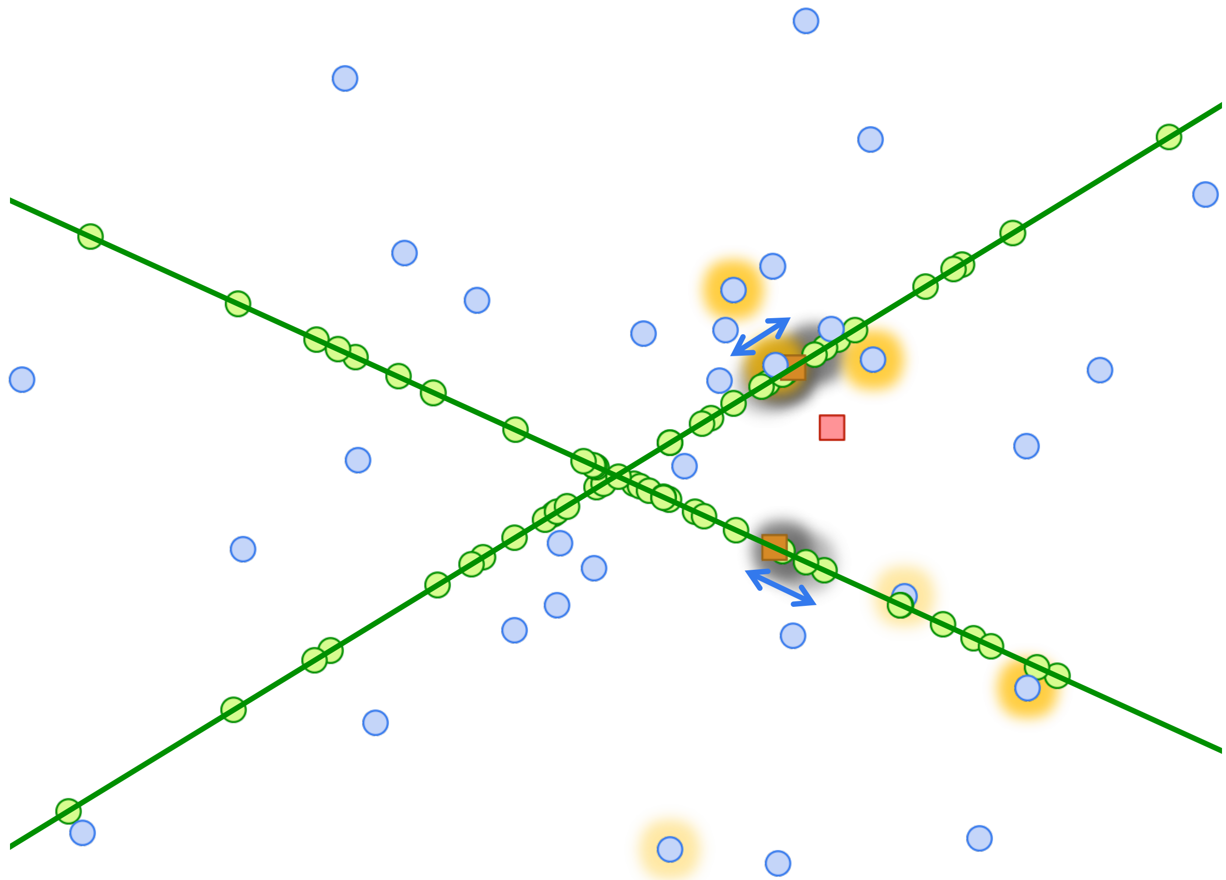
Compare projected distances of points on the frontier and visit the one with the shortest projected distance.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI



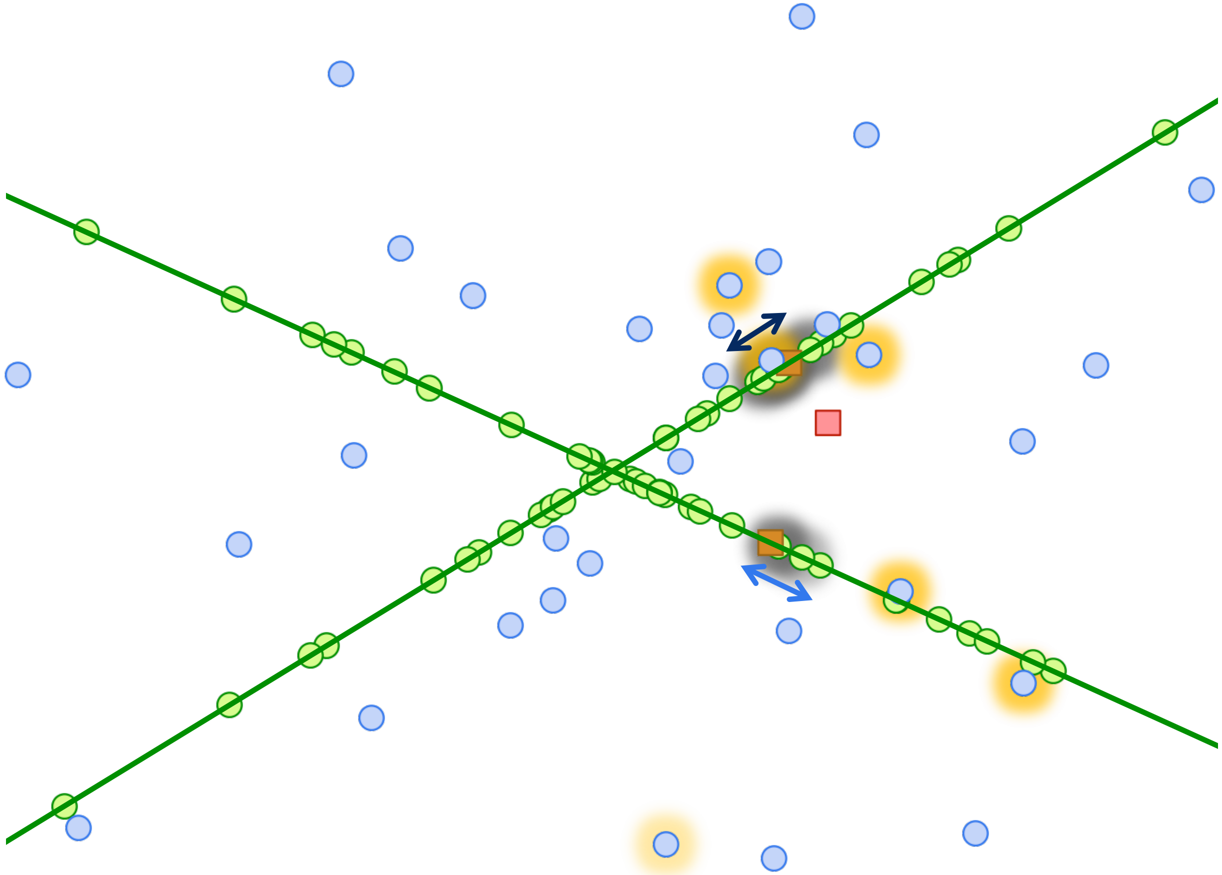
Find the next closest point along the projection direction that has just been processed and add it to the frontier.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI



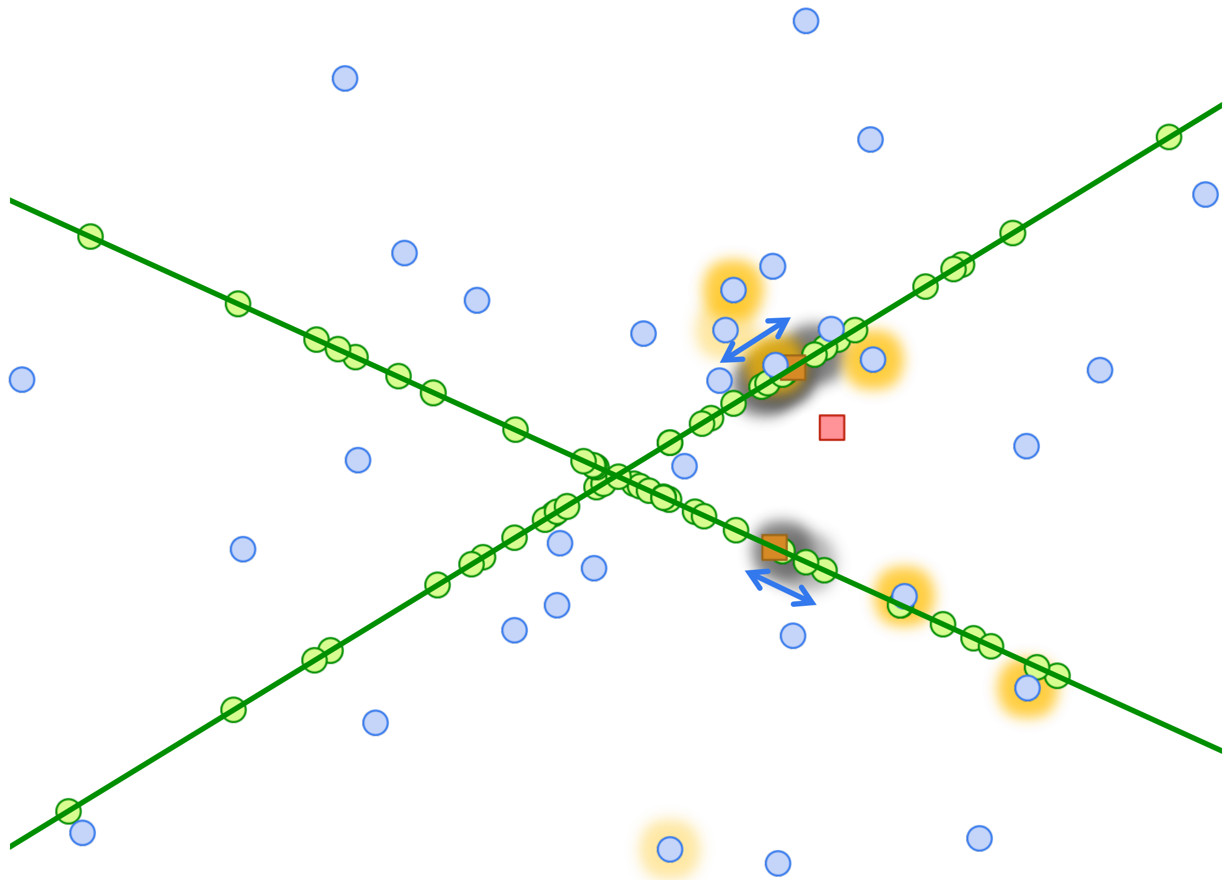
Compare projected distances of points on the frontier and visit the one with the shortest projected distance.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI



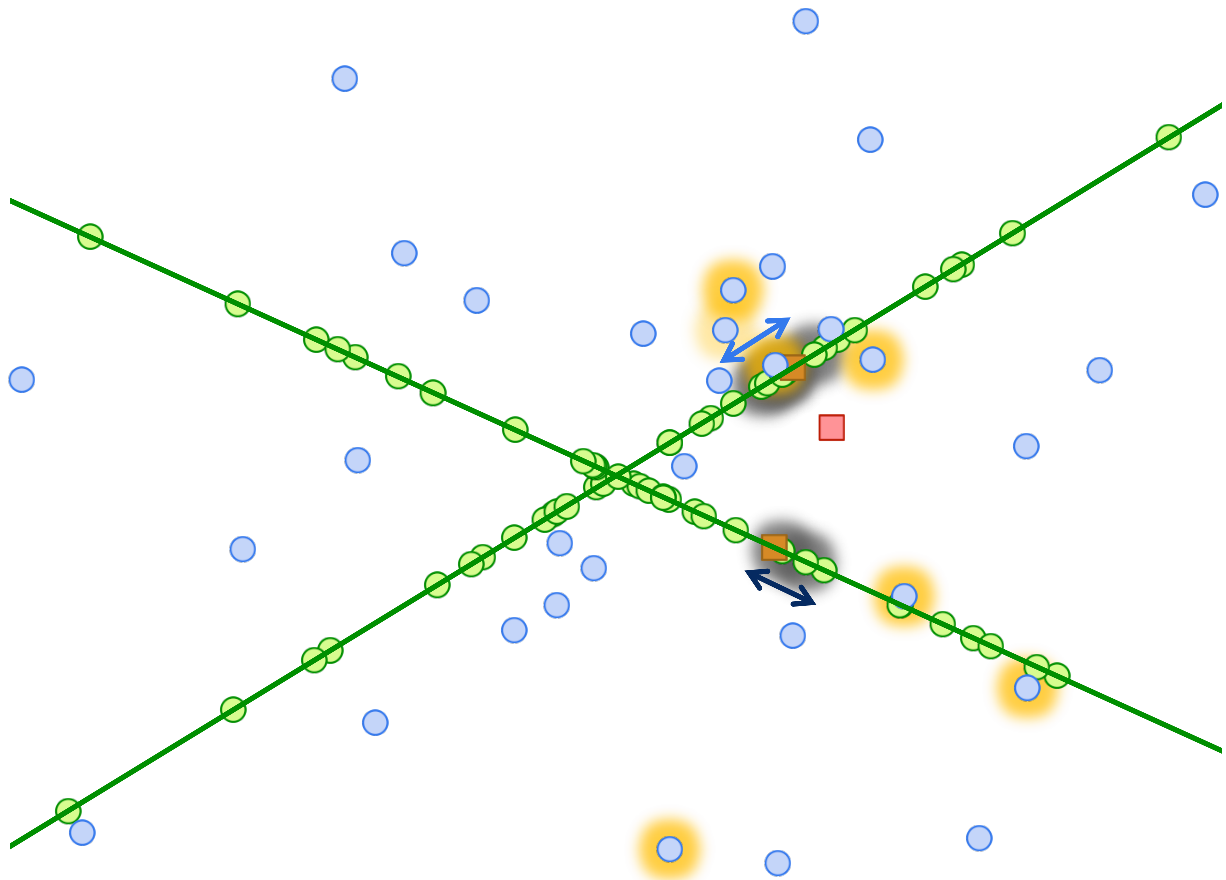
Find the next closest point along the projection direction that has just been processed and add it to the frontier.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI



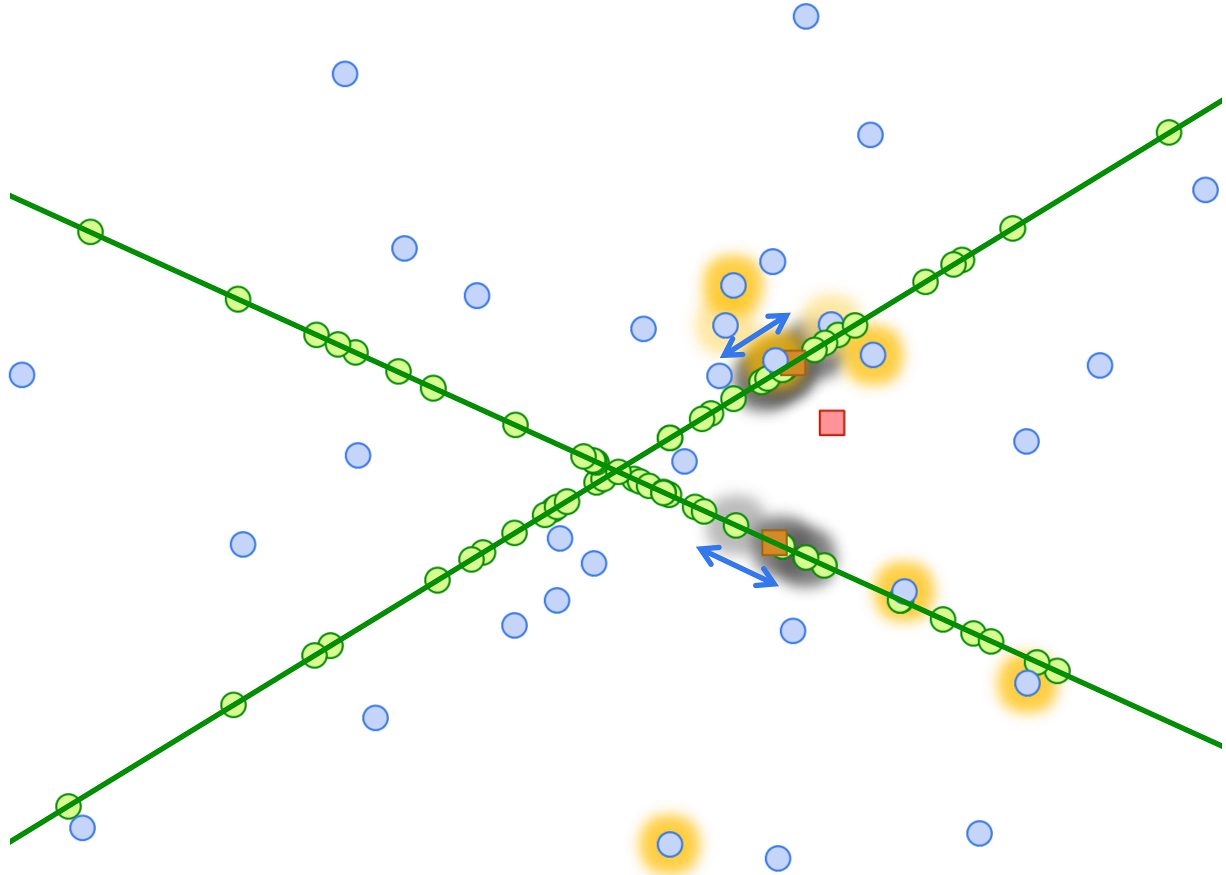
Compare projected distances of points on the frontier and visit the one with the shortest projected distance.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI



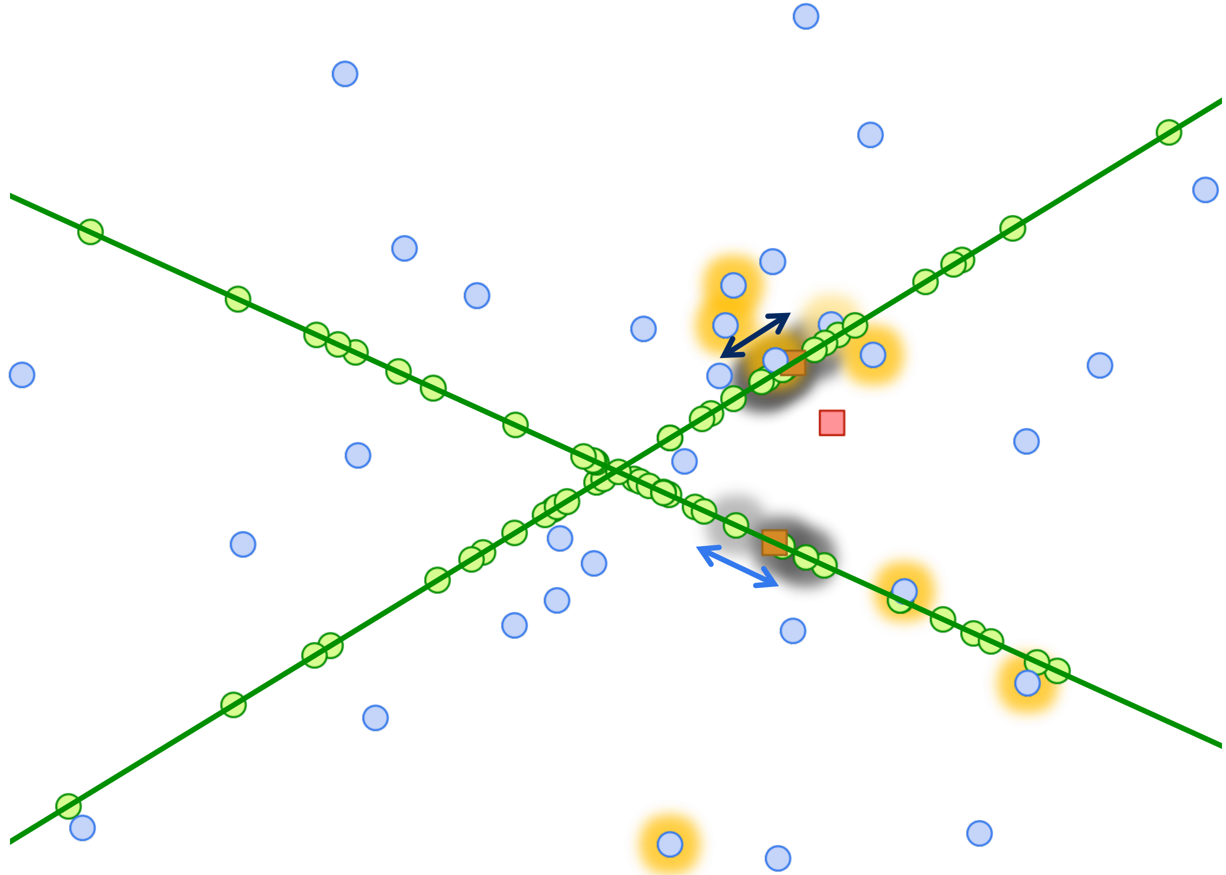
Find the next closest point along the projection direction that has just been processed and add it to the frontier.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

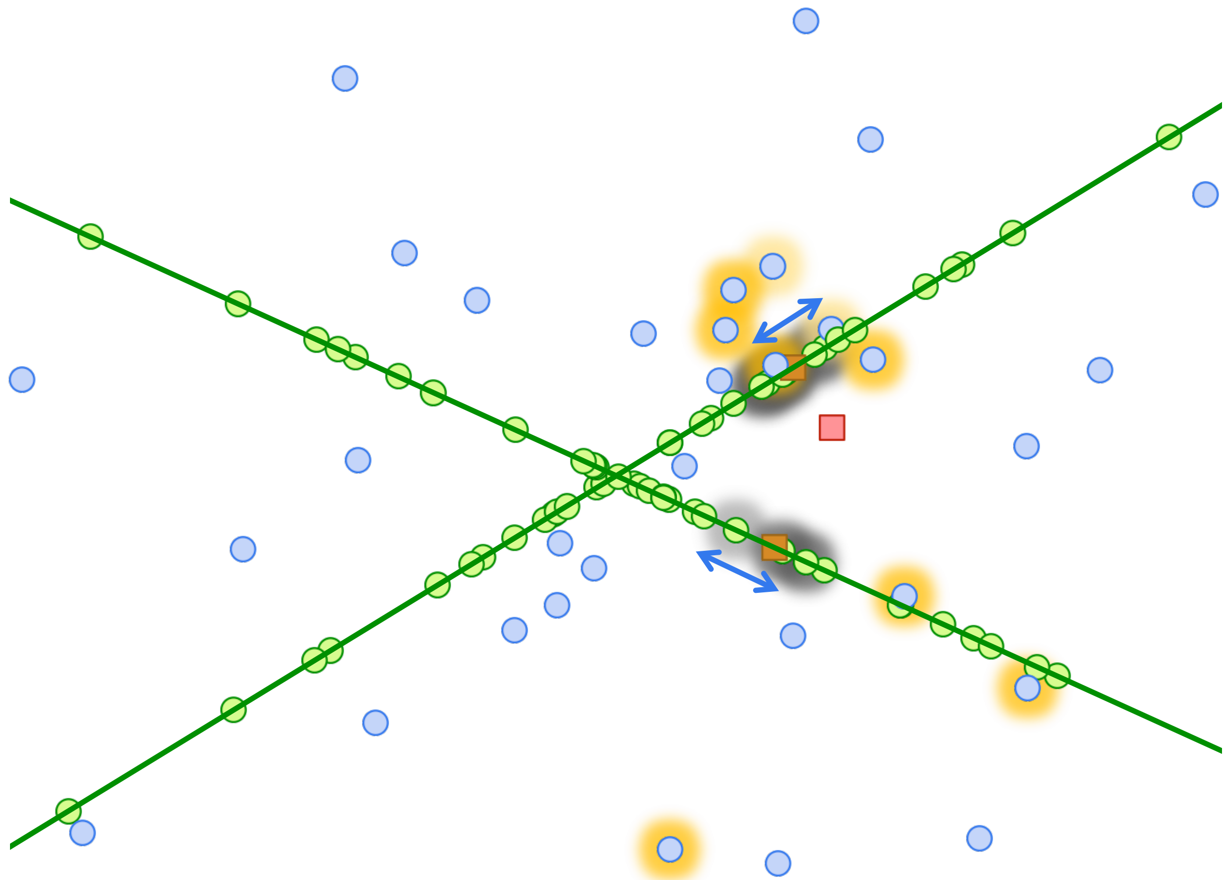
# Prioritized DCI



Compare projected distances of points on the frontier and visit the one with the shortest projected distance.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# Prioritized DCI



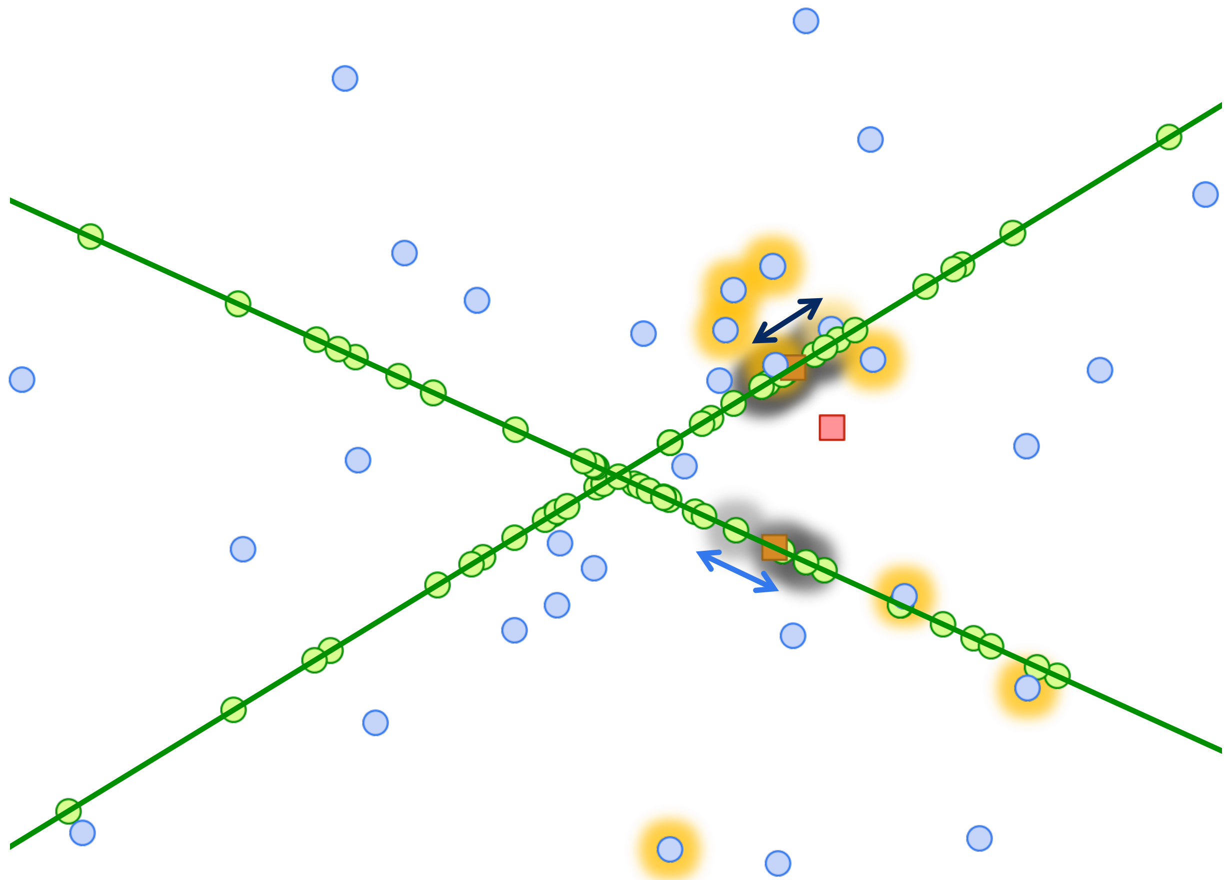
Find the next closest point along the projection direction that has just been processed and add it to the frontier.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI



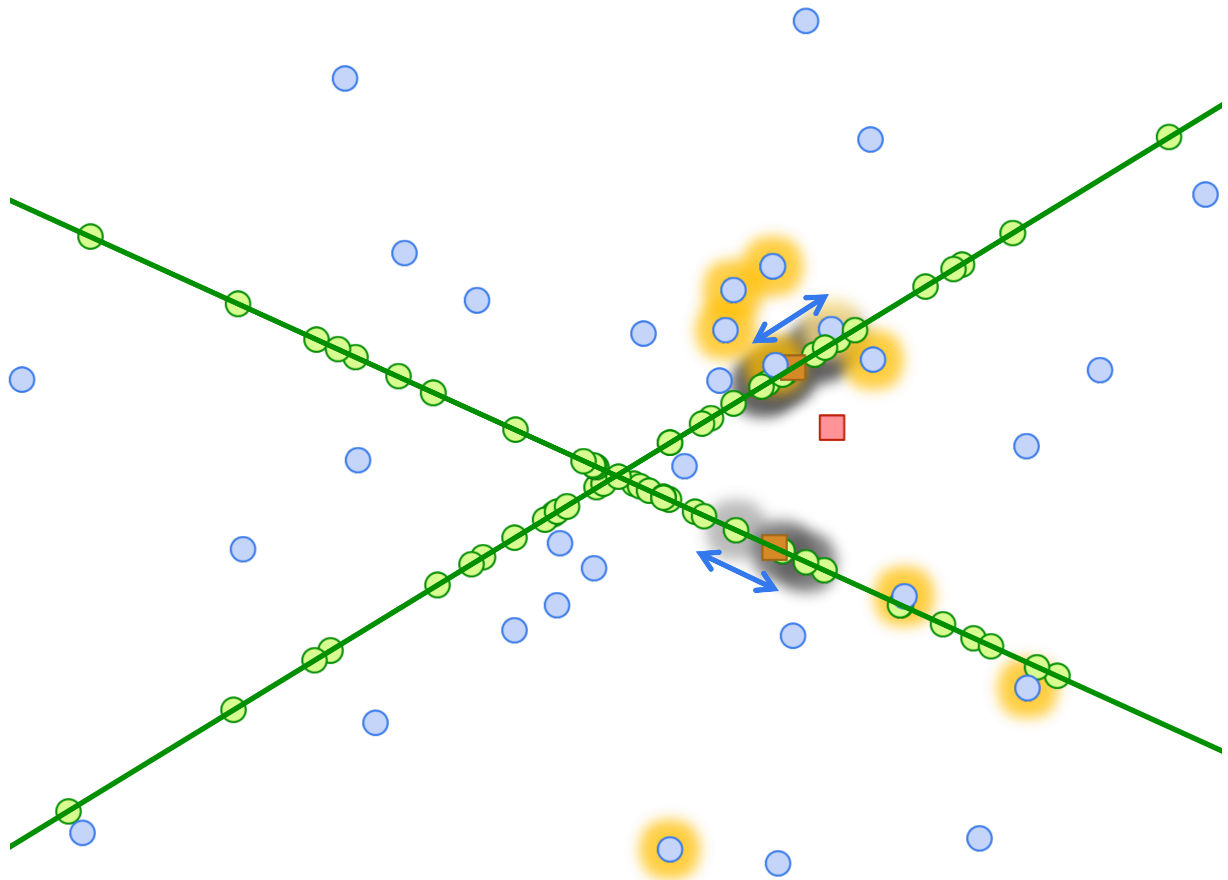
Compare projected distances of points on the frontier and visit the one with the shortest projected distance.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI



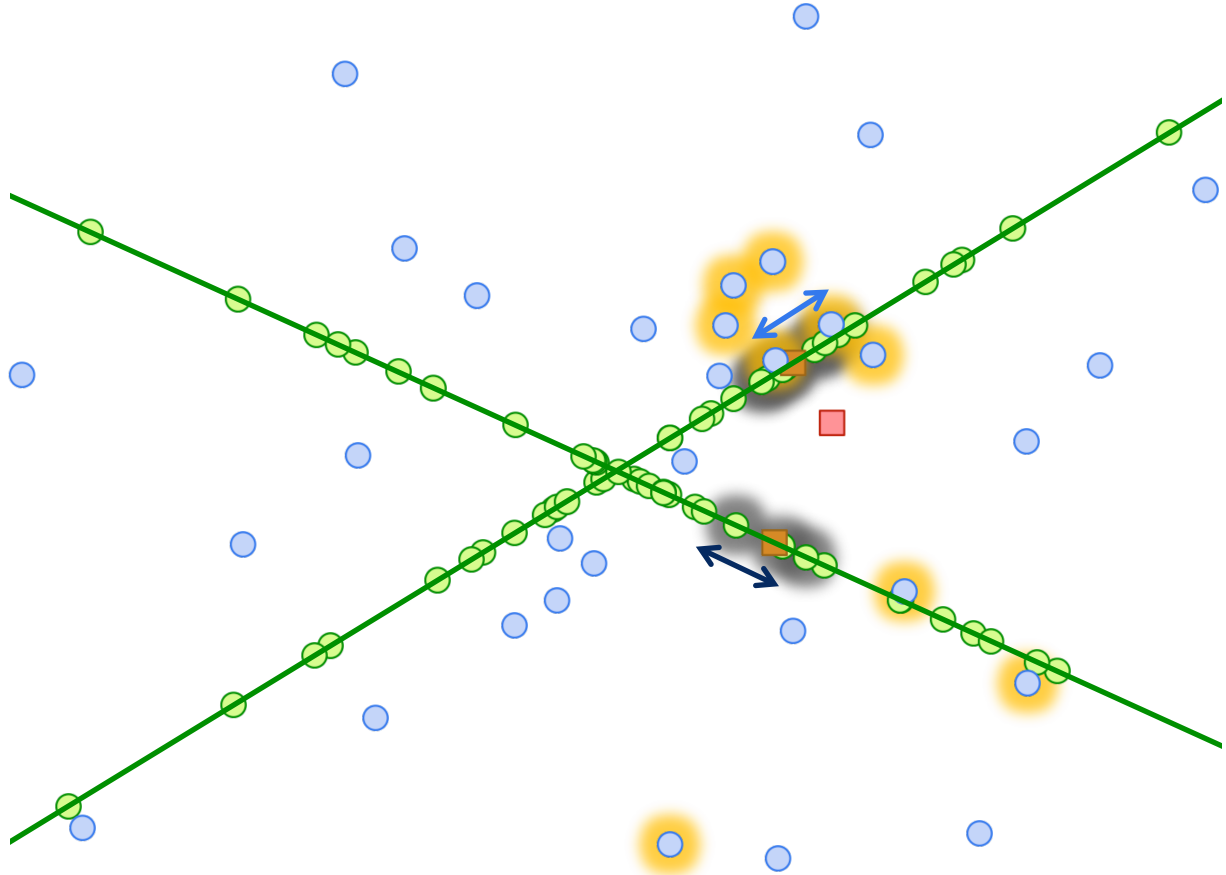
Find the next closest point along the projection direction that has just been processed and add it to the frontier.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

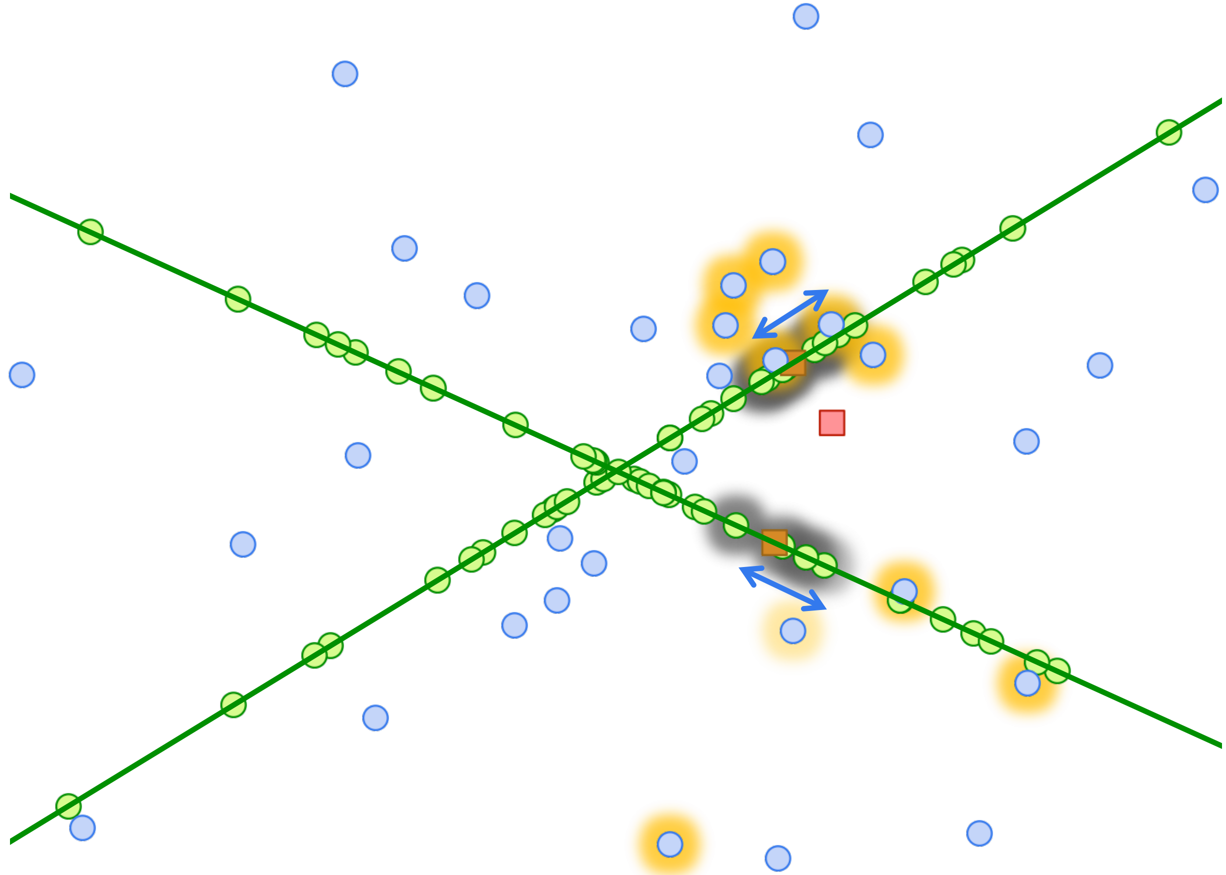
# Prioritized DCI



Compare projected distances of points on the frontier and visit the one with the shortest projected distance.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

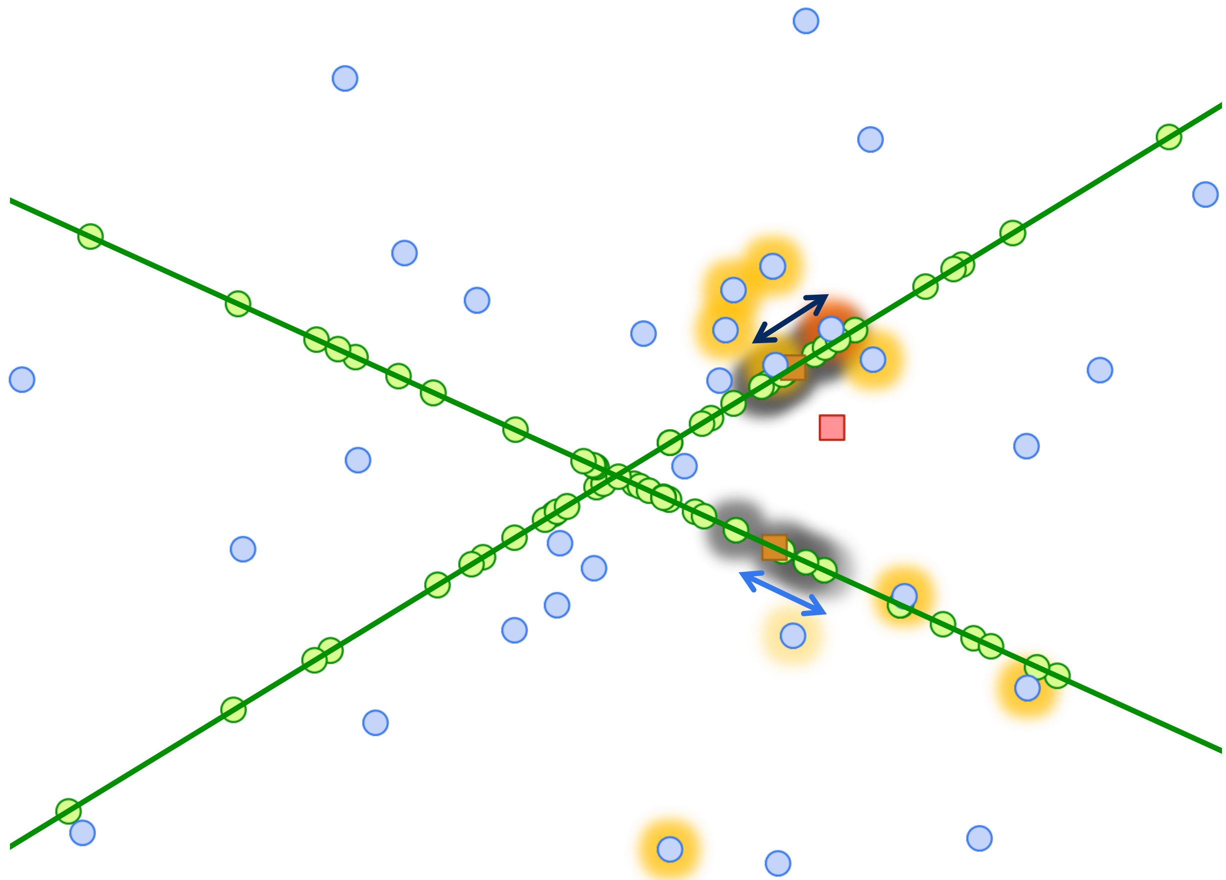
# Prioritized DCI



Find the next closest point along the projection direction that has just been processed and add it to the frontier.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# Prioritized DCI



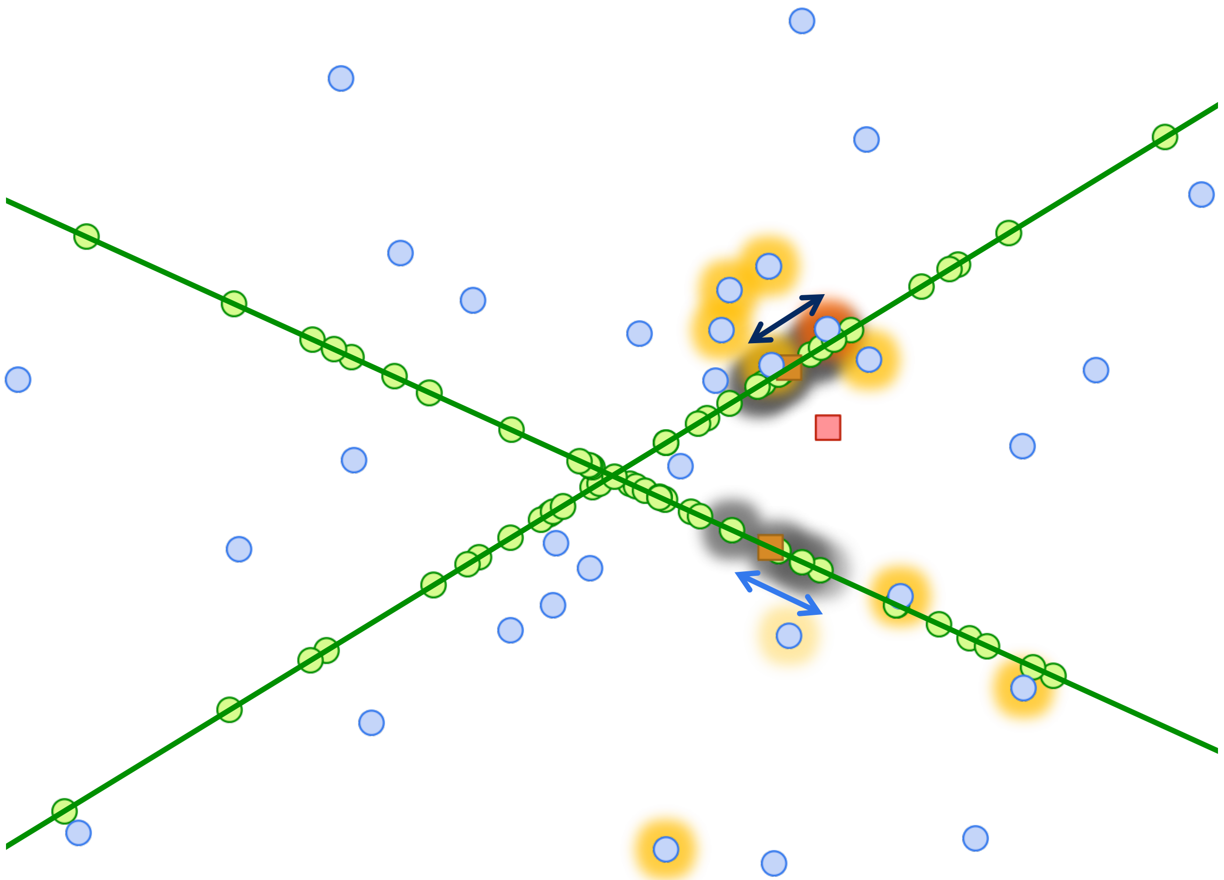
Compare projected distances of points on the frontier and visit the one with the shortest projected distance.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI



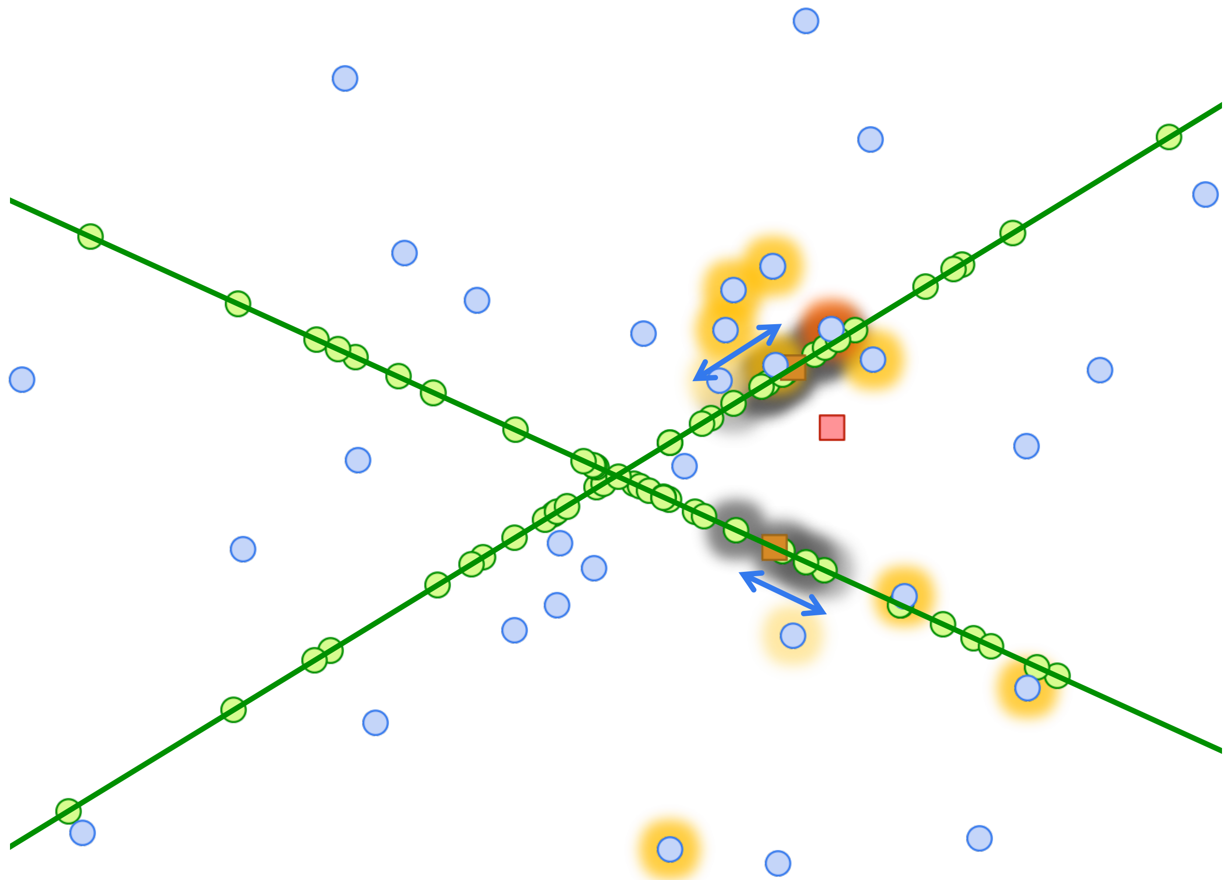
There is now a point that has been visited along all projection directions. We add it to the candidate set.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI



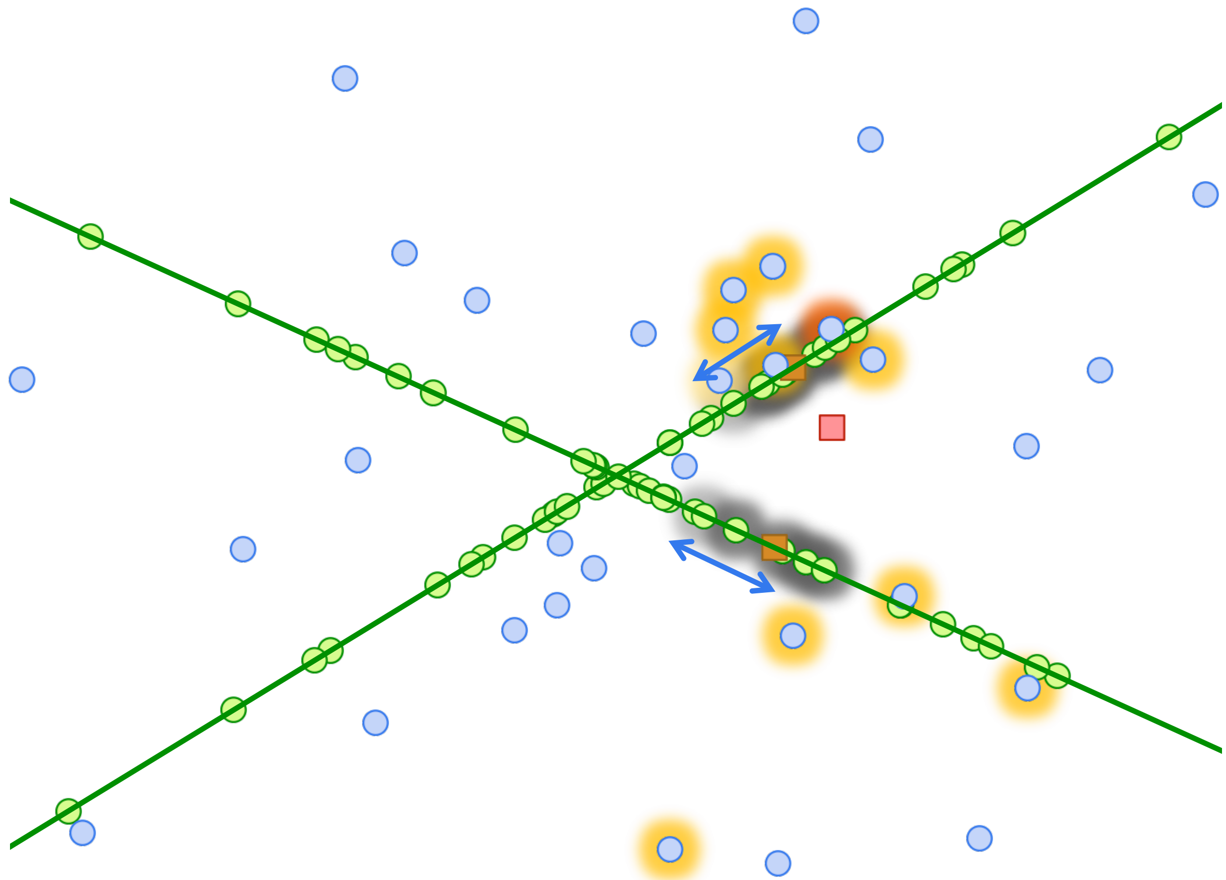
Visit the next point.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Prioritized DCI



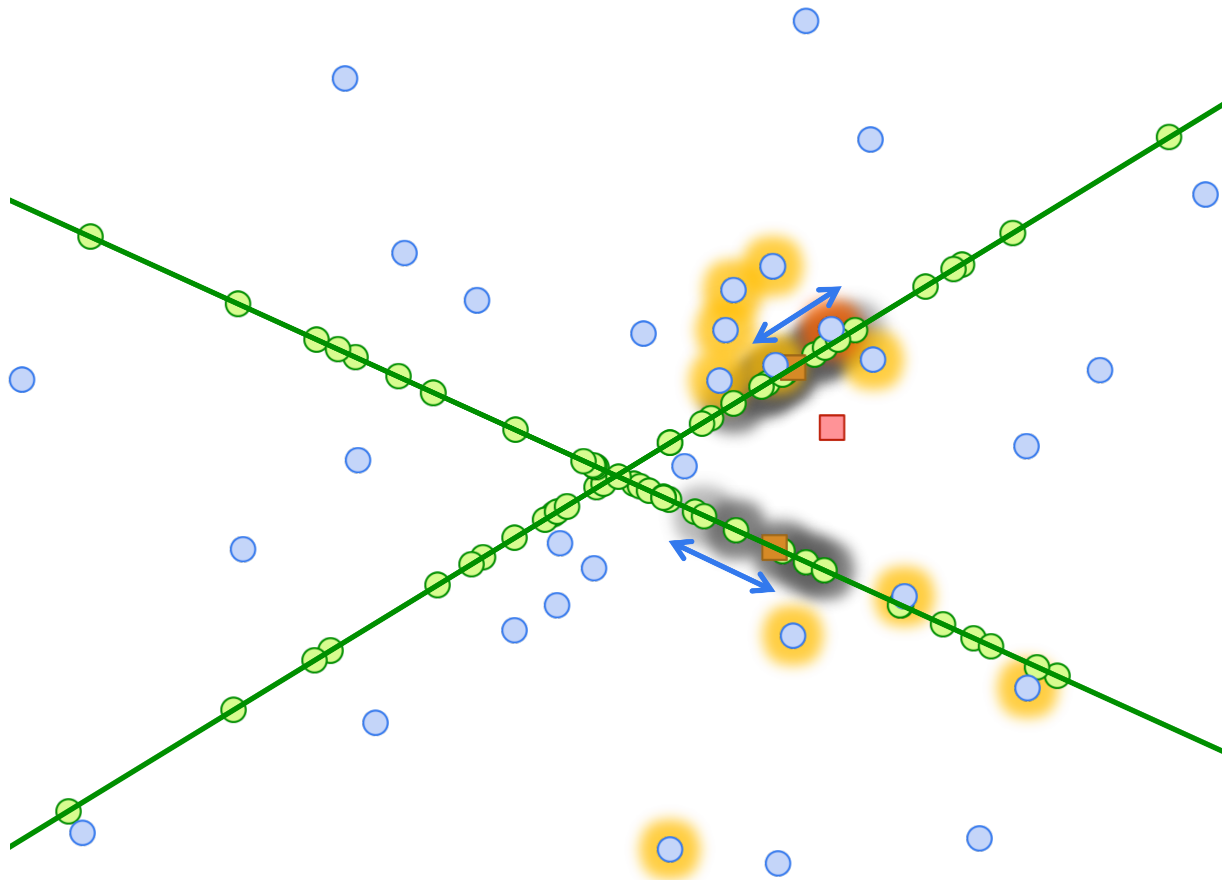
Visit the next point.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

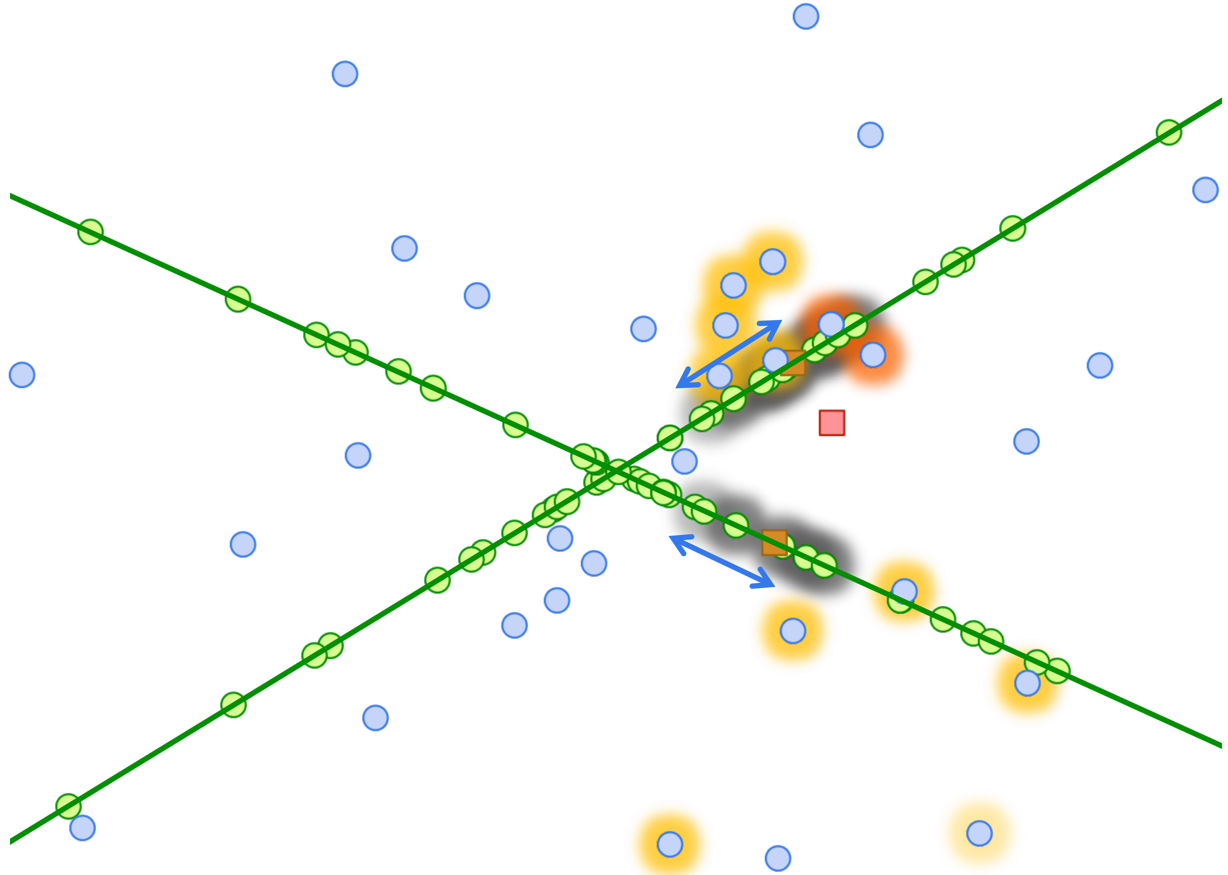
# Prioritized DCI



Visit the next point.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# Prioritized DCI



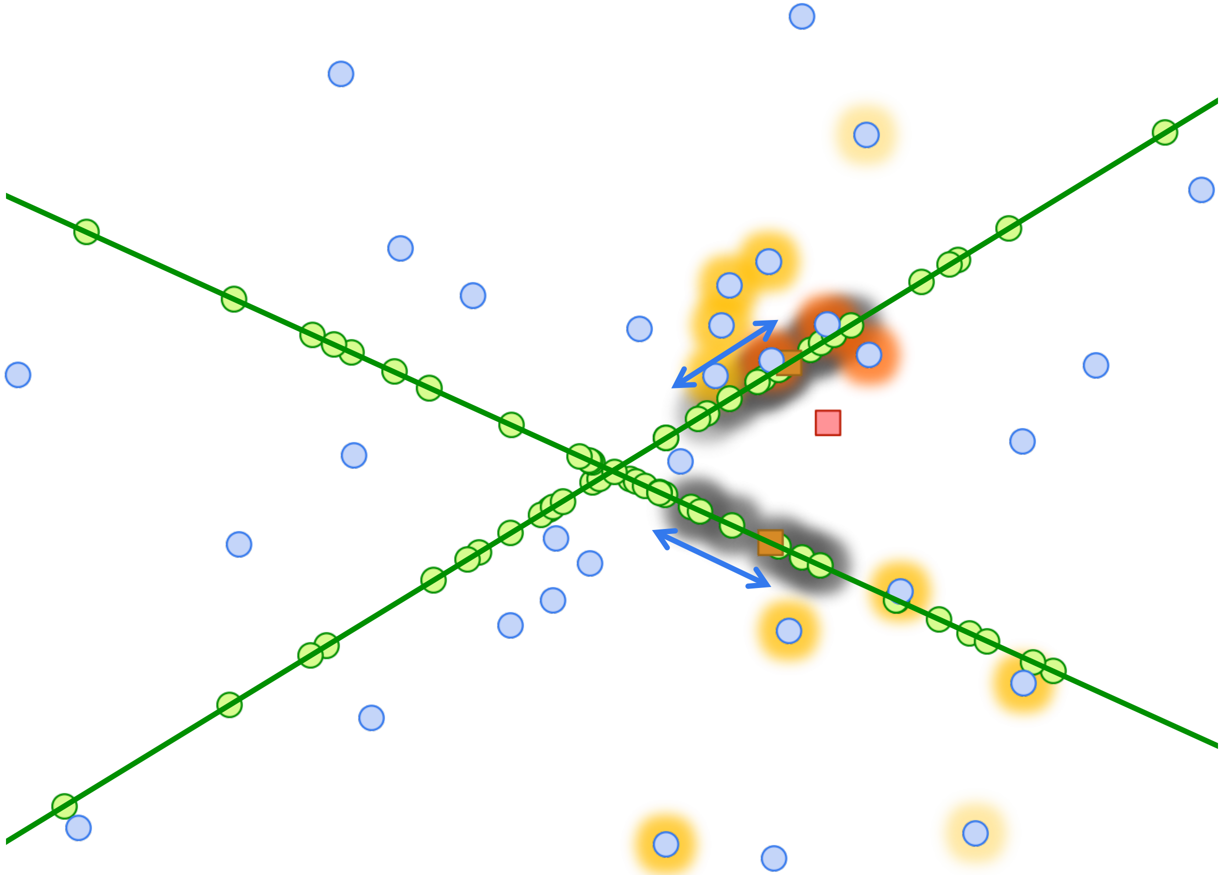
Visit the next point and add it to the candidate set.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

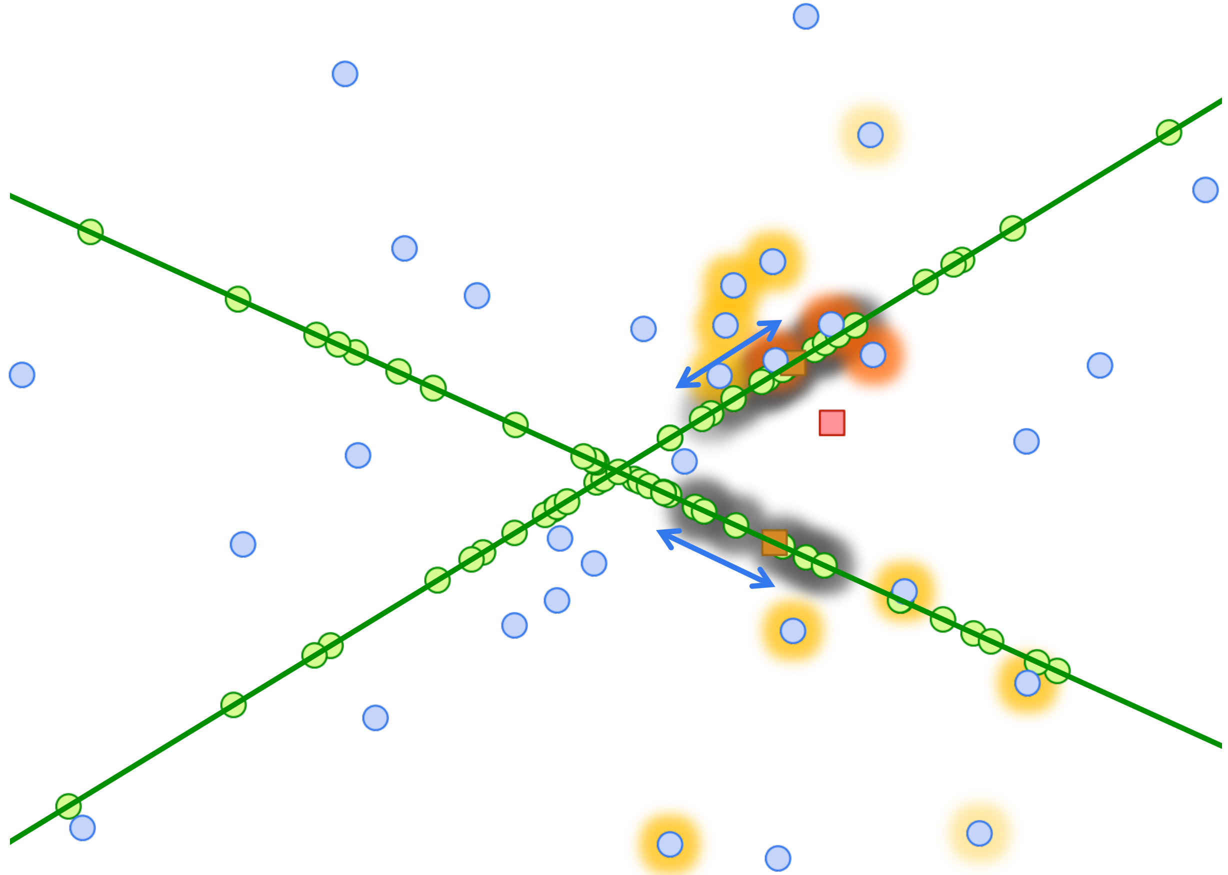
# Prioritized DCI



Visit the next point and add it to the candidate set.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# Prioritized DCI



Perform exhaustive search over candidate points and return *k* points that are closest to the query.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# Intuition

- Points are added to the candidate set in the order of their maximum projected distance to the query.
- Maximum projected distance is a lower bound on the true distance.
- As the number of projection directions increases, this lower bound approaches the true distance.

$$\max_j \left\{ \left| \langle p^i, u_j \rangle - \langle q, u_j \rangle \right| \right\} = \max_j \left\{ \left| \langle p^i - q, u_j \rangle \right| \right\} \leq \left\| p^i - q \right\|_2$$

where $\left\| u_j \right\|_2 = 1 \quad \forall j$

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Lemma

- We derived the following lemma, which may be of independent interest:
- For any set of events $E_1, \ldots, E_n$, the probability that at least $m$ of them occur is at most:

$$\frac{1}{m} \sum_{i=1}^{n} \Pr\left(E_i\right)$$

- When $m = 1$, this reduces to the union bound.

(Proof is in the "Fast k-Nearest Neighbour Search via Prioritized DCI" paper, though two students, Eric Xia and Zipeng Qin, later came up with simpler proofs, one using measure theory, and one using Markov's inequality.)

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# Complexity

- Construction Time: $O(m(dn + n \log n))$
- Query Time: $O(dk \max(\log(n/k), (n/k)^{1-m/d'})+$
  $mk \log m(\max(\log(n/k), (n/k)^{1-1/d'})))$
- Insertion Time: $O(m(d + \log n))$
- Deletion Time: $O(m \log n)$
- Space: $O(mn)$

where $m \geq 1$ is the number of projection directions chosen by the user.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Complexity

- Construction Time: $O(m(dn + n\log n))$
- Query Time: $O(\boxed{d}k \max(\log(n/k), \boxed{(n/k)^{1-m/d'}})+$
  $mk \log m(\max(\log(n/k), \boxed{(n/k)^{1-1/d'}})))$
- Insertion ... $\log n))$
- Deletion ... $)$
- Space: $O(mn)$

*Linear* dependence on ambient dimensionality

*Sublinear* dependence on intrinsic dimensionality

where $m \geq 1$ is the number of projection directions chosen by the user.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Complexity

- Construction Time: $O(m(dn + n \log n))$
- Query Time: $O(dk \max(\log(n/k), \boxed{(n/k)^{1-m/d'}})+$
  $mk \log m(\max(\log(n/k), (n/k)^{1-1/d'})))$
- Insertion Time: $O(m(d + \log n))$
- Deletion Time: $O(m \log n)$
- Space: $O(mn)$

A linear increase in intrinsic dimensionality can be mostly counteracted with a linear increase in the number of projection directions.

where $m \geq 1$ is the number of projection directions chosen by the user.

K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



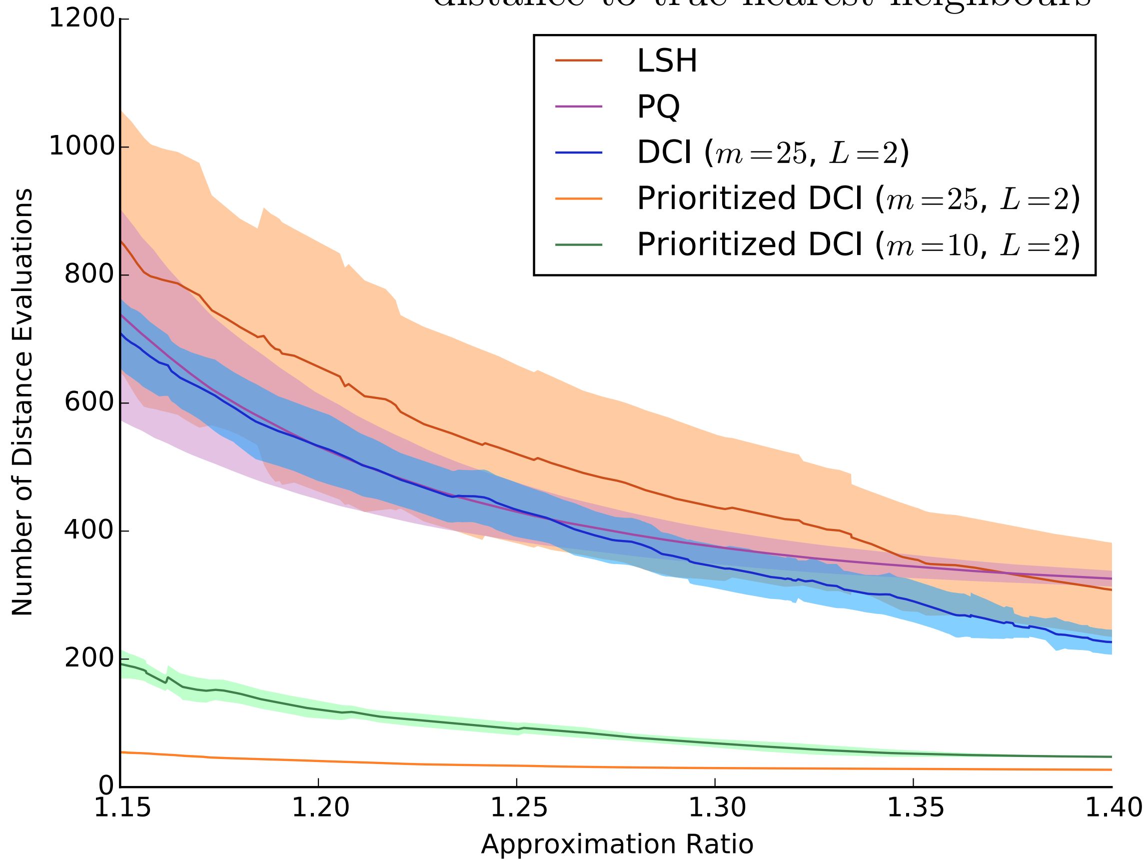
Berkeley
UNIVERSITY OF CALIFORNIA

# Experiments

# Query Time on CIFAR-100

$$\text{approximation ratio} = \frac{\text{distance to retrieved nearest neighbours}}{\text{distance to true nearest neighbours}}$$

14x

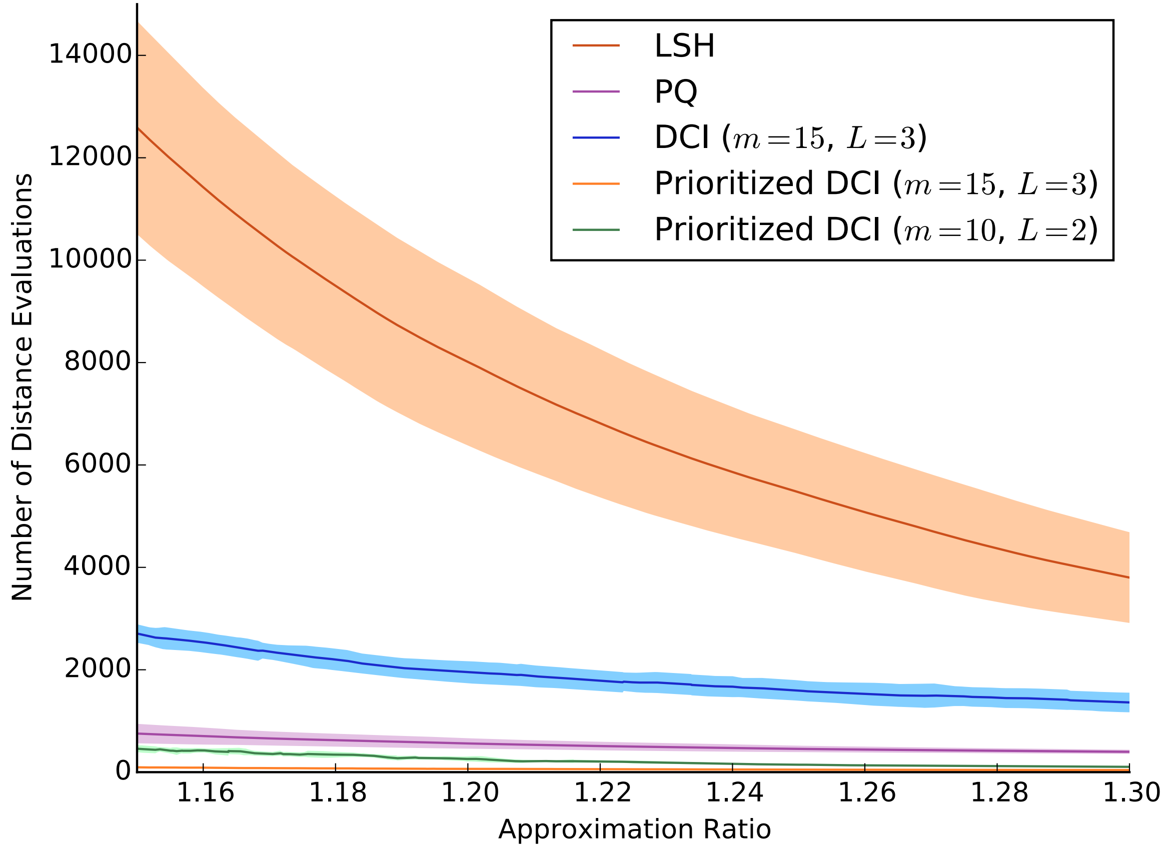



K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# Query Time on MNIST

$$\text{approximation ratio} = \frac{\text{distance to retrieved nearest neighbours}}{\text{distance to true nearest neighbours}}$$

116x




K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

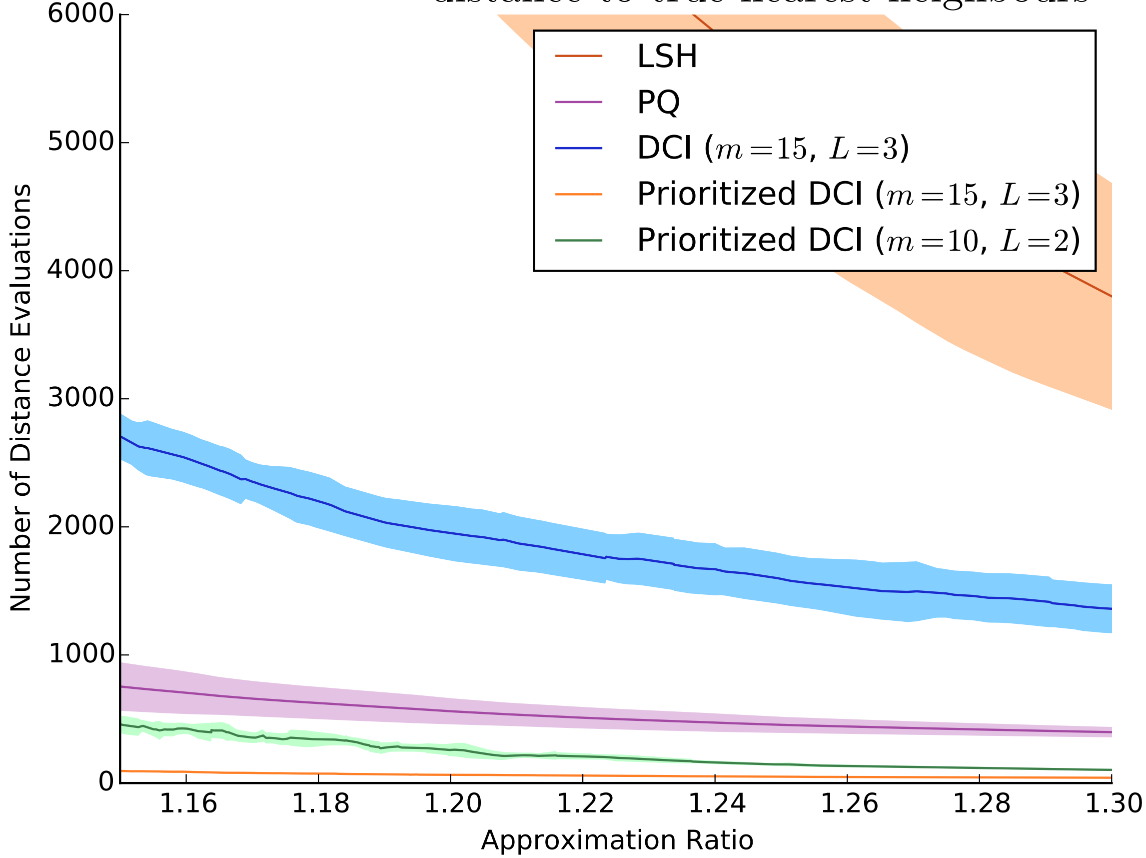


Berkeley
UNIVERSITY OF CALIFORNIA

# Query Time on MNIST

$$\text{approximation ratio} = \frac{\text{distance to retrieved nearest neighbours}}{\text{distance to true nearest neighbours}}$$

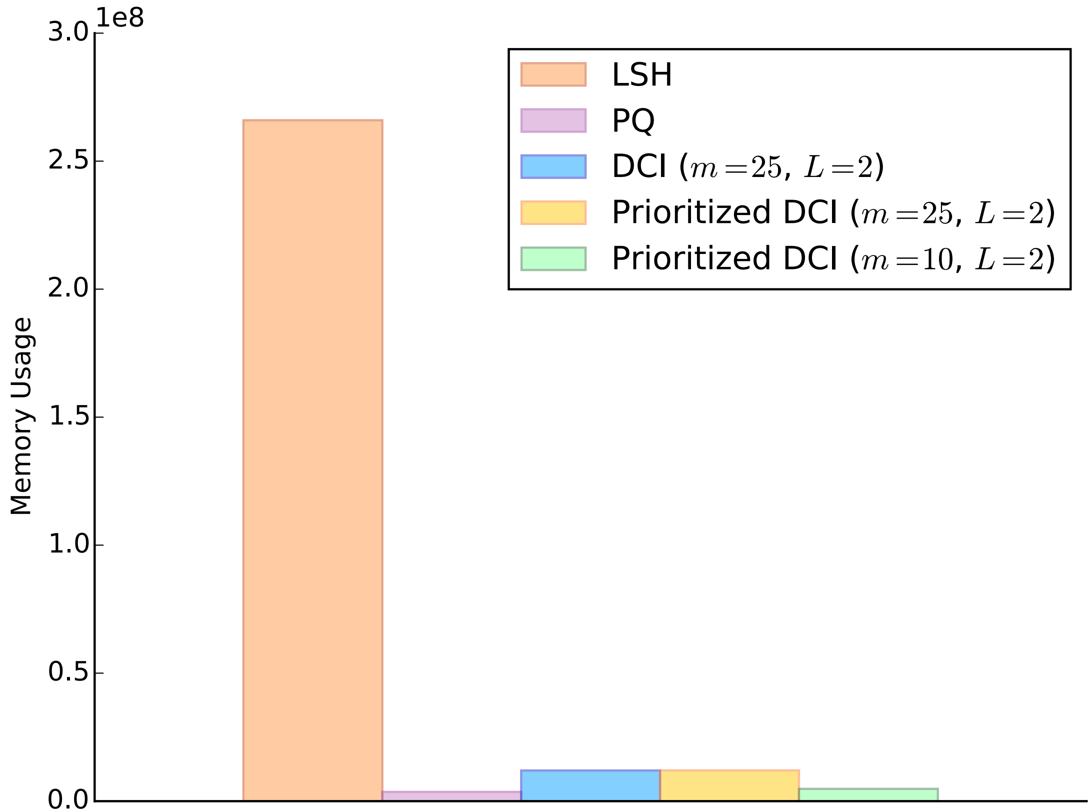
116x




K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# Space Efficiency on CIFAR-100

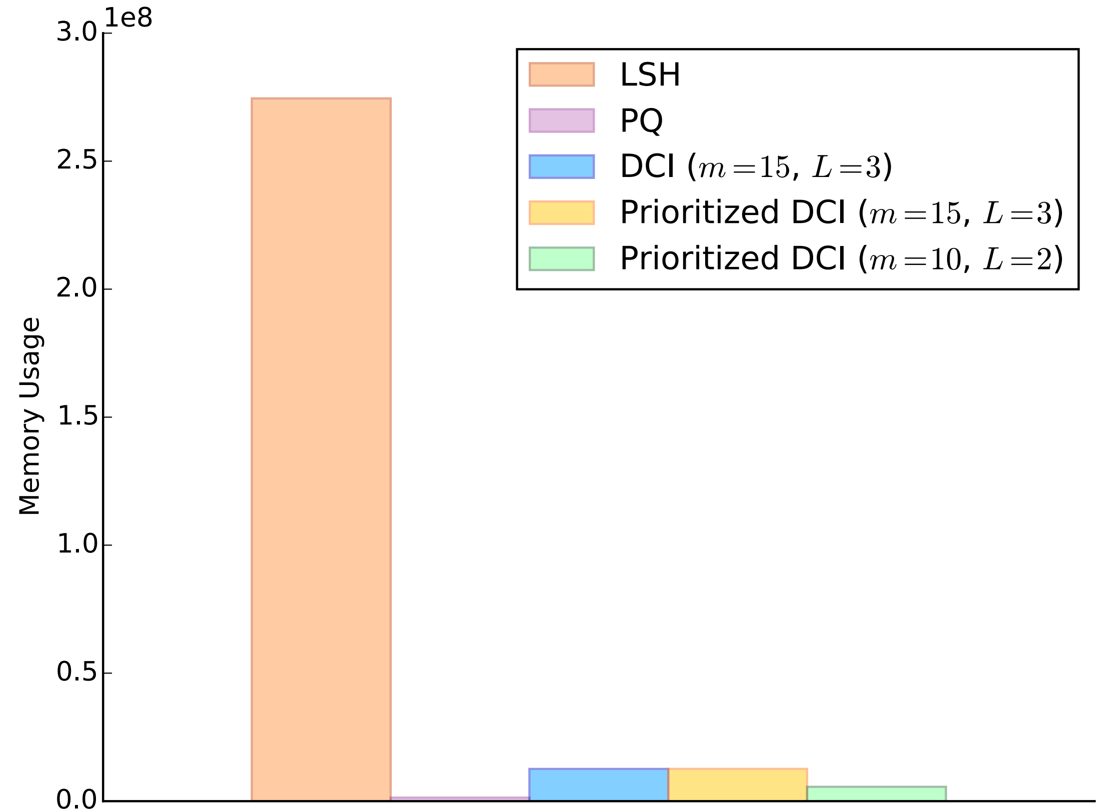
22x




K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017

# Space Efficiency on MNIST

21x




K. Li and J. Malik. Fast *k*-nearest neighbour search via Prioritized DCI. *ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA

# Agenda

1. Implicit Maximum Likelihood Estimation
2. Comparison to Generative Adversarial Nets (GANs)
3. Why Maximum Likelihood
4. Equivalence to Maximum Likelihood
5. Fast Nearest Neighbour Search
6. *Applications to Conditional Image Synthesis*

K. Li and J. Malik. Implicit Maximum Likelihood Estimation. *arXiv: 1809.09087*, 2018



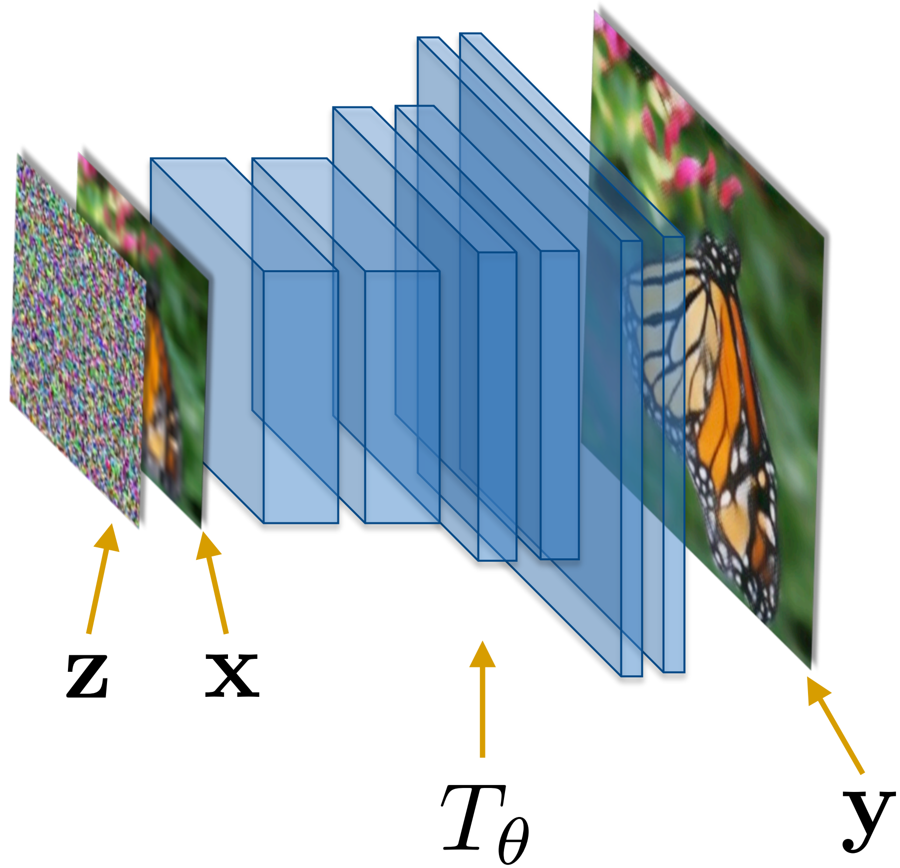
Berkeley
UNIVERSITY OF CALIFORNIA

# Application to Conditional Image Synthesis

# Conditional Implicit Probabilistic Model

$$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

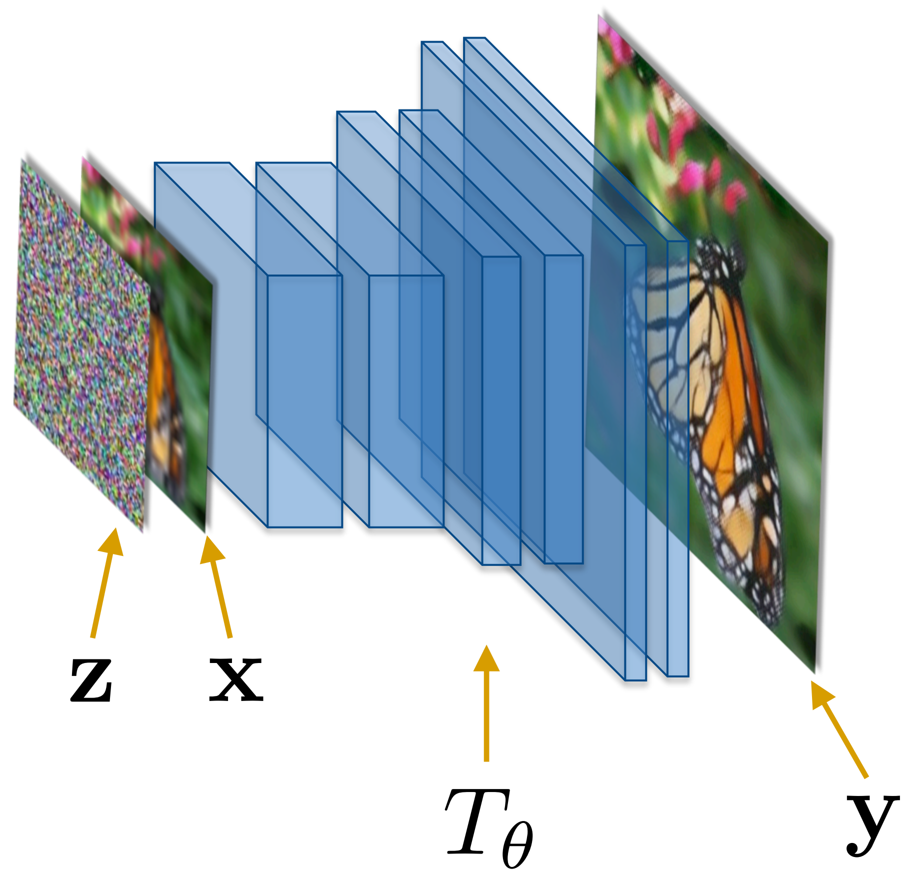
$$\mathbf{y} = T_\theta(\mathbf{x}, \mathbf{z})$$



# Conditional Implicit Probabilistic Model

$$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$\mathbf{y} = T_\theta(\mathbf{x}, \mathbf{z})$$



# Conditional IMLE

- Let $\tilde{\mathbf{y}}_{i,1}^{\theta}, \ldots, \tilde{\mathbf{y}}_{i,m}^{\theta}$ denote i.i.d. samples from $p(\tilde{\mathbf{y}}|\mathbf{x}_i;\theta)$.
- To train the model, we solve the following optimization problem:

$$\hat{\theta}_{\mathrm{IMLE}} := \arg\min_{\theta} \mathbb{E}_{\left\{\tilde{\mathbf{y}}_{i,j}^{\theta}\right\}_{i\in[n],j\in[m]}} \left[\sum_{i=1}^{n} \min_{j\in\{1,\ldots,m\}} \left\|\phi(\tilde{\mathbf{y}}_{i,j}^{\theta}) - \phi(\mathbf{y}_i)\right\|_2^2\right]$$

# Ill-Posed Problems

- Modelling the full distribution of outputs given the input is important for ill-posed problems.
- Many image synthesis problems are of this flavour, since there are usually many plausible output images that are consistent with the input.



Berkeley
UNIVERSITY OF CALIFORNIA

# Multimodal Super-Resolution

Input

Output 1

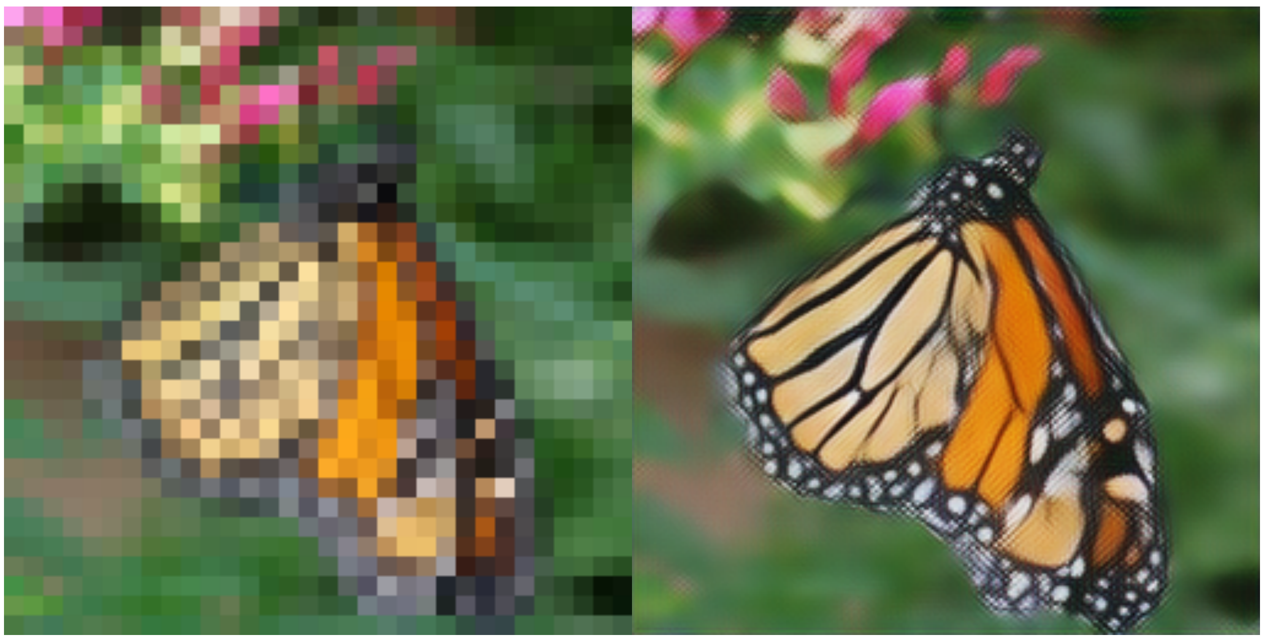
K. Li*, S. Peng* and J. Malik. Super-Resolution via Conditional Implicit Maximum Likelihood Estimation. *arXiv: 1810.01406*, 2018

UNIVERSITY OF TORONTO

Berkeley UNIVERSITY OF CALIFORNIA

# Multimodal Super-Resolution

Input Output 2

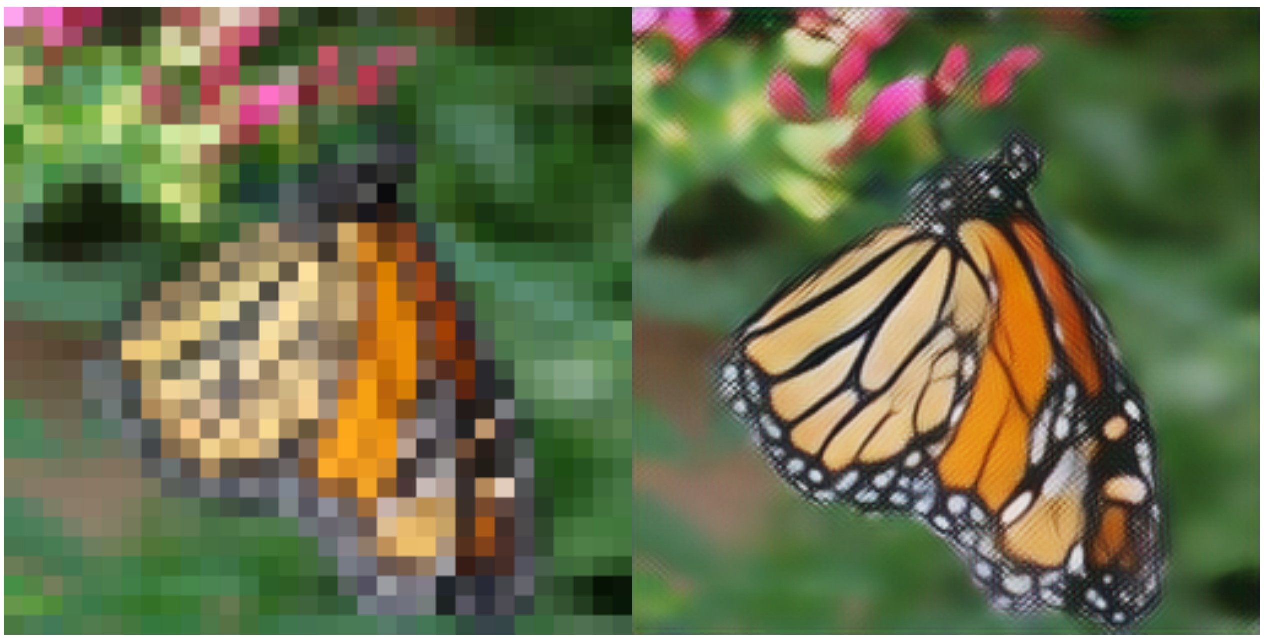


K. Li*, S. Peng* and J. Malik. Super-Resolution via Conditional Implicit Maximum Likelihood Estimation. *arXiv: 1810.01406*, 2018

UNIVERSITY OF TORONTO Berkeley UNIVERSITY OF CALIFORNIA

# Multimodal Super-Resolution

Input Output 3

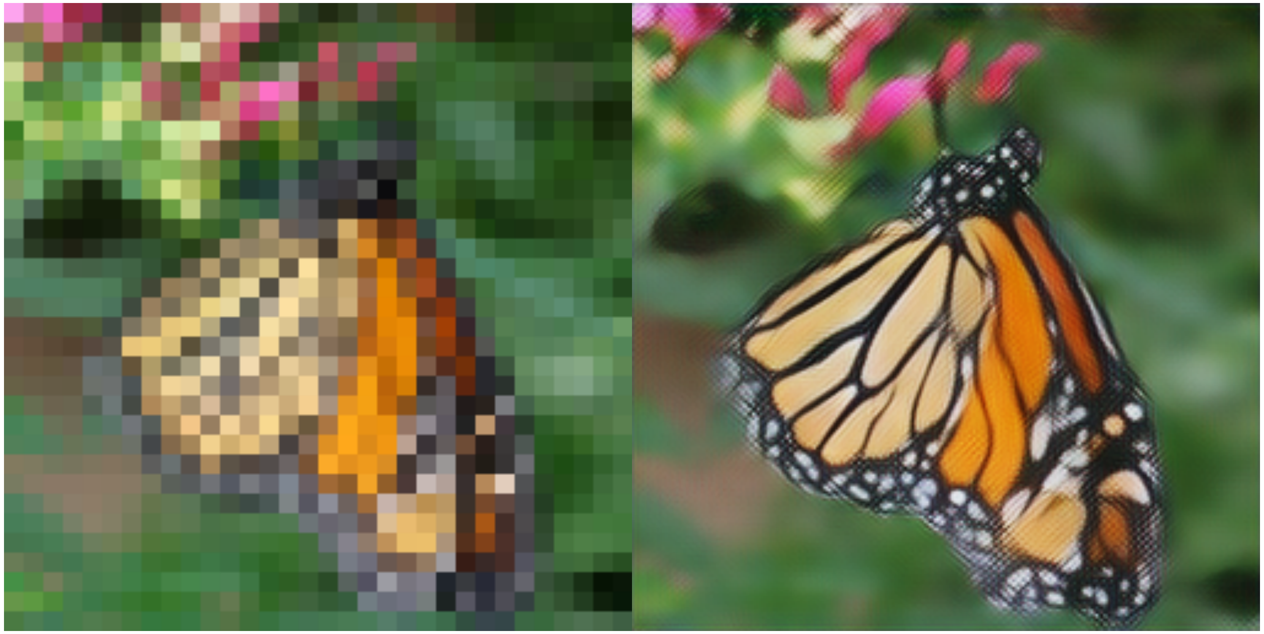


K. Li*, S. Peng* and J. Malik. Super-Resolution via Conditional Implicit Maximum Likelihood Estimation. *arXiv: 1810.01406*, 2018

UNIVERSITY OF TORONTO Berkeley UNIVERSITY OF CALIFORNIA

# Multimodal Super-Resolution

Input | Output 4

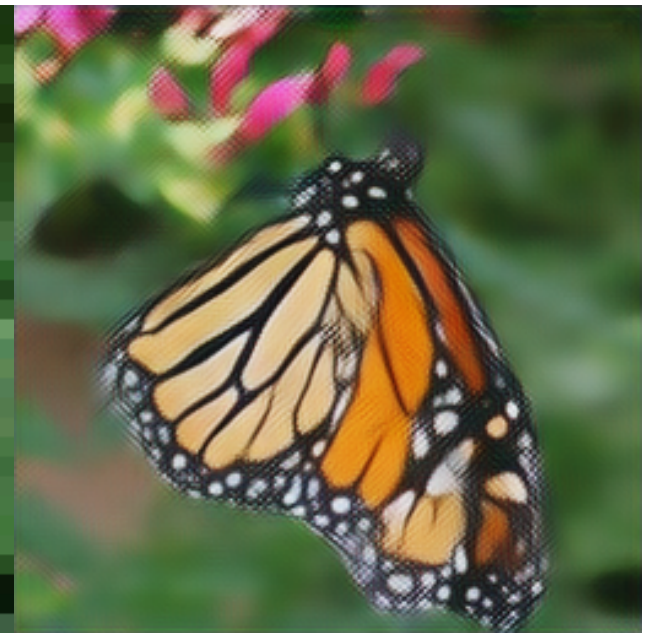


K. Li*, S. Peng* and J. Malik. Super-Resolution via Conditional Implicit Maximum Likelihood Estimation. *arXiv: 1810.01406*, 2018

# Multimodal Super-Resolution

Input

Output 5



K. Li*, S. Peng* and J. Malik. Super-Resolution via Conditional Implicit Maximum Likelihood Estimation. *arXiv: 1810.01406*, 2018

UNIVERSITY OF TORONTO

Berkeley
UNIVERSITY OF CALIFORNIA

# Multimodal Image Synthesis from Semantic Layout

Input

K. Li*, T. Zhang* and J. Malik. Diverse Image Synthesis from Semantic Layouts via Conditional IMLE. *arXiv:1811.12373*, 2018

# Multimodal Image Synthesis from Semantic Layout

Predicted Output 1

K. Li*, T. Zhang* and J. Malik. Diverse Image Synthesis from Semantic Layouts via Conditional IMLE. *arXiv:1811.12373*, 2018

南京大學 Berkeley UNIVERSITY OF CALIFORNIA

# Multimodal Image Synthesis from Semantic Layout

Predicted Output 2

K. Li*, T. Zhang* and J. Malik. Diverse Image Synthesis from Semantic Layouts via Conditional IMLE. *arXiv:1811.12373*, 2018

南京大學 Berkeley UNIVERSITY OF CALIFORNIA

# Multimodal Image Synthesis from Semantic Layout

Predicted Output 3

K. Li*, T. Zhang* and J. Malik. Diverse Image Synthesis from Semantic Layouts via Conditional IMLE. *arXiv:1811.12373*, 2018

# Multimodal Image Synthesis from Semantic Layout

Predicted Output 4

K. Li*, T. Zhang* and J. Malik. Diverse Image Synthesis from Semantic Layouts via Conditional IMLE. *arXiv:1811.12373*, 2018

南京大學 Berkeley UNIVERSITY OF CALIFORNIA

# Multimodal Image Synthesis from Semantic Layout

Predicted Output 5

K. Li*, T. Zhang* and J. Malik. Diverse Image Synthesis from Semantic Layouts via Conditional IMLE. *arXiv:1811.12373*, 2018

南京大學 Berkeley UNIVERSITY OF CALIFORNIA

# Multimodal Image Synthesis from Semantic Layout

Predicted Output 6

K. Li*, T. Zhang* and J. Malik. Diverse Image Synthesis from Semantic Layouts via Conditional IMLE. *arXiv:1811.12373*, 2018

# Multimodal Image Synthesis from Semantic Layout

Predicted Output 7

K. Li*, T. Zhang* and J. Malik. Diverse Image Synthesis from Semantic Layouts via Conditional IMLE. *arXiv:1811.12373*, 2018

# Multimodal Image Synthesis from Semantic Layout

Predicted Output 8

K. Li*, T. Zhang* and J. Malik. Diverse Image Synthesis from Semantic Layouts via Conditional IMLE. *arXiv:1811.12373*, 2018

南京大學 Berkeley UNIVERSITY OF CALIFORNIA

# Multimodal Image Synthesis from Semantic Layout

Predicted Output 9

K. Li*, T. Zhang* and J. Malik. Diverse Image Synthesis from Semantic Layouts via Conditional IMLE. *arXiv:1811.12373*, 2018

南京大學 Berkeley UNIVERSITY OF CALIFORNIA

# Multimodal Image Synthesis from Semantic Layout

Predicted Output 10

K. Li*, T. Zhang* and J. Malik. Diverse Image Synthesis from Semantic Layouts via Conditional IMLE. *arXiv:1811.12373*, 2018

南京大學 Berkeley UNIVERSITY OF CALIFORNIA

# For More Details…

**Implicit Maximum Likelihood Estimation**
Ke Li, Jitendra Malik
*arXiv:1809.09087*, 2018

**Super-Resolution via Conditional Implicit Maximum Likelihood Estimation**
Ke Li*, Shichong Peng*, Jitendra Malik
*arXiv:1810.01406*, 2018

**Diverse Image Synthesis from Semantic Layouts via Conditional IMLE**
Ke Li*, Tianhao Zhang*, Jitendra Malik
*arXiv:1811.12373*, 2018

**Fast *k*-Nearest Neighbour Search via Dynamic Continuous Indexing**
Ke Li, Jitendra Malik
*ICML*, 2016

**Fast *k*-Nearest Neighbour Search via Prioritized DCI**
Ke Li, Jitendra Malik
*ICML*, 2017



Berkeley
UNIVERSITY OF CALIFORNIA