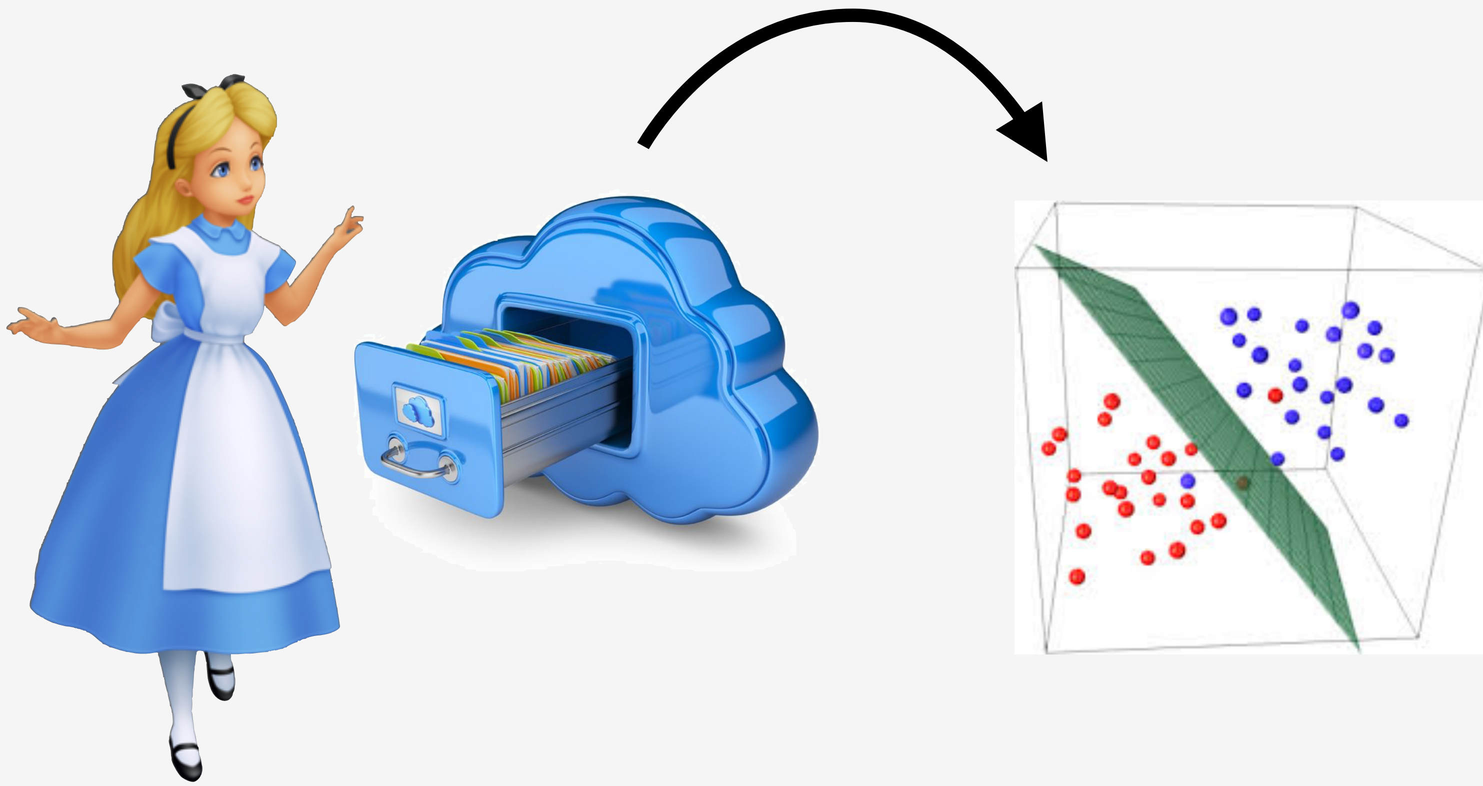
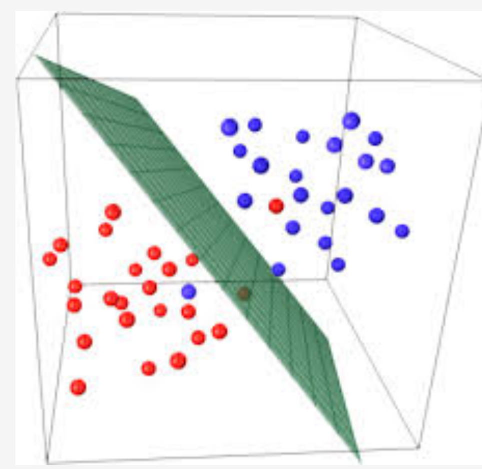
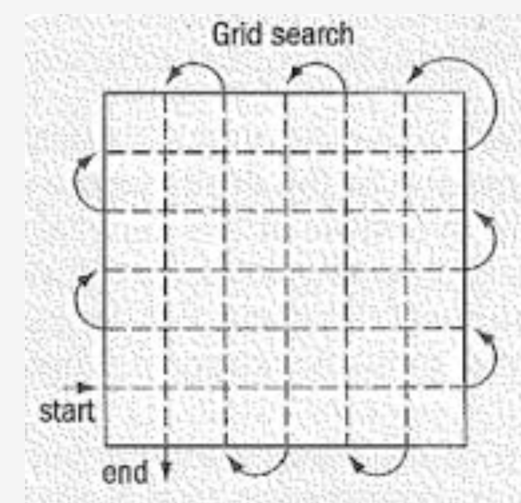
Differentially Private Bayesian Optimization

Matt J. Kusner*
Roman Garnett
Jake Gardner
Kilian Weinberger

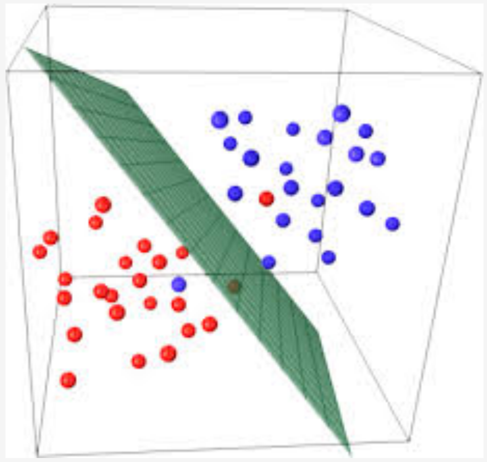
*part of work done while at author was at Yahoo! Labs

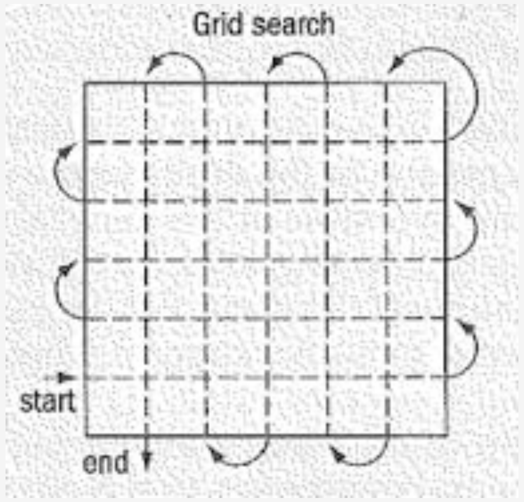





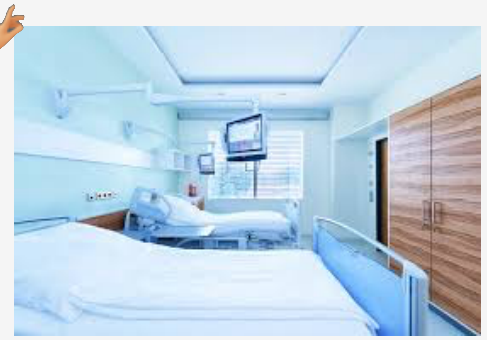

Grid search
start
end

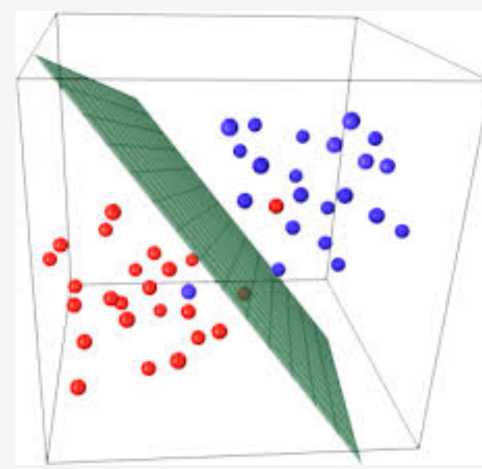
Goal: Release best validation accuracy and/or best hyperparameters

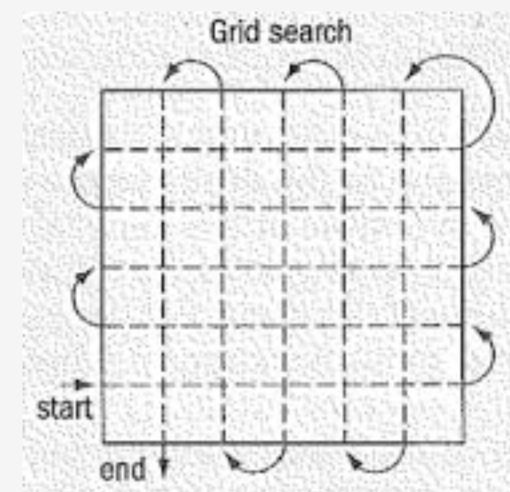




Goal: Release best validation accuracy and/or best hyperparameters



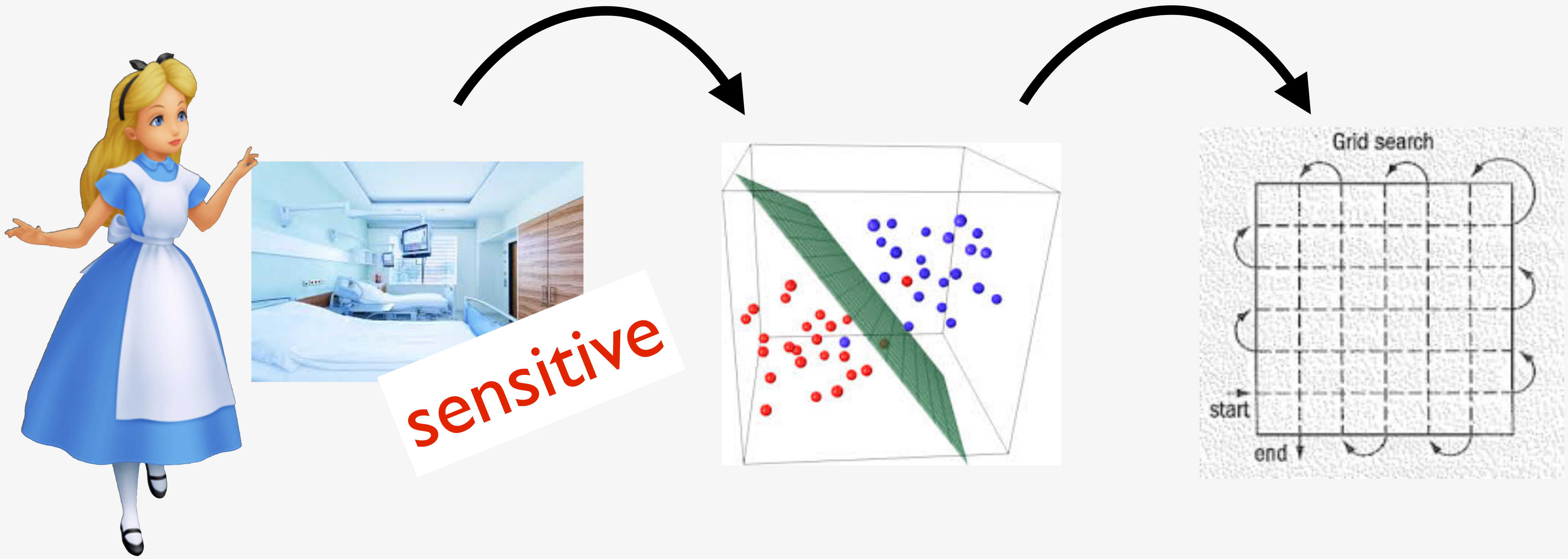




"I can detect cancer with 98% accuracy"

"These are the model hyperparameter values: [0.51,0.87]"

Goal: Release best validation accuracy and/or best hyperparameters

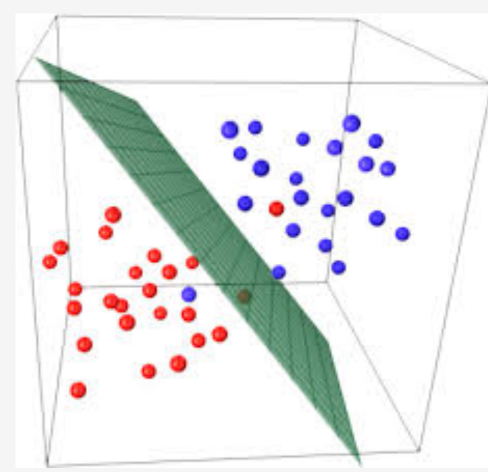


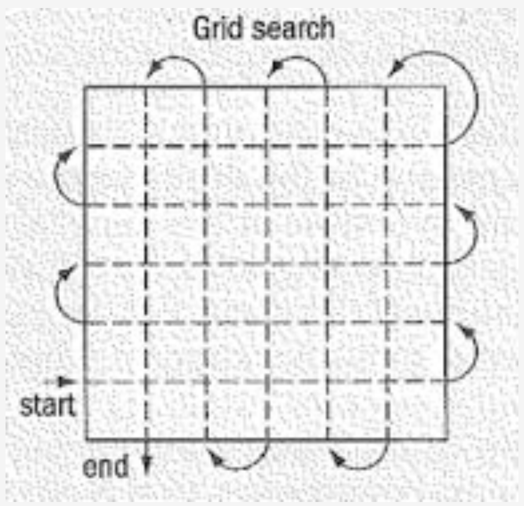
"I can detect cancer with 98% accuracy"

"These are the model hyperparameter values: [0.51,0.87]"

Goal: Release best validation accuracy and/or best hyperparameters

sensitive

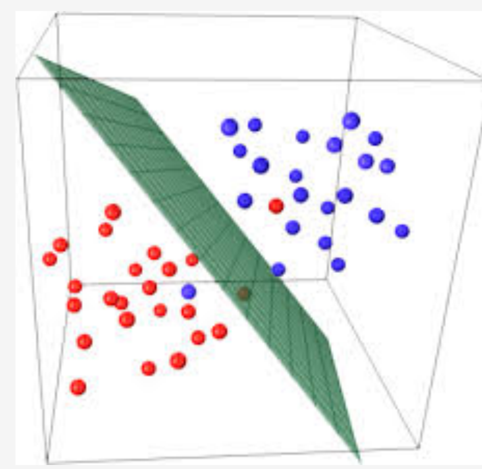


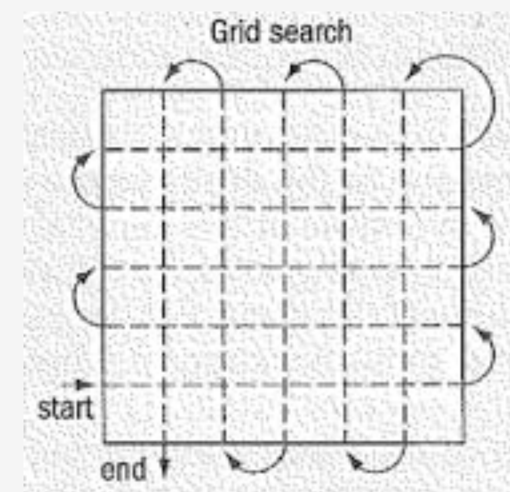


Problem: Releasing hyperparameters from grid search compromises privacy

[Chaudhuri & Vinterbo, 2013]

Goal: Release best validation accuracy and/or best hyperparameters

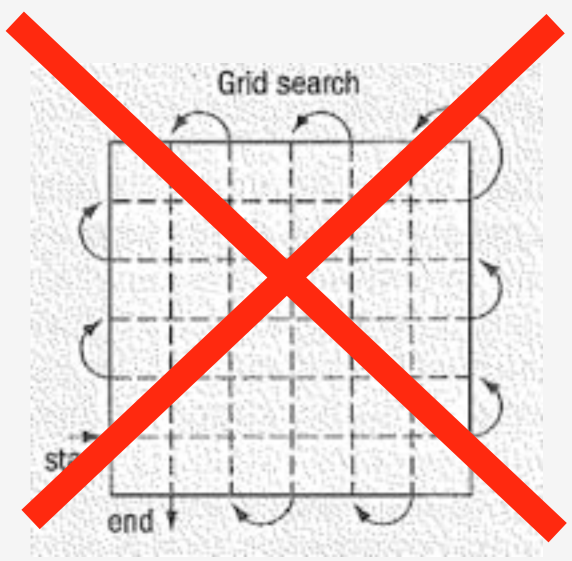




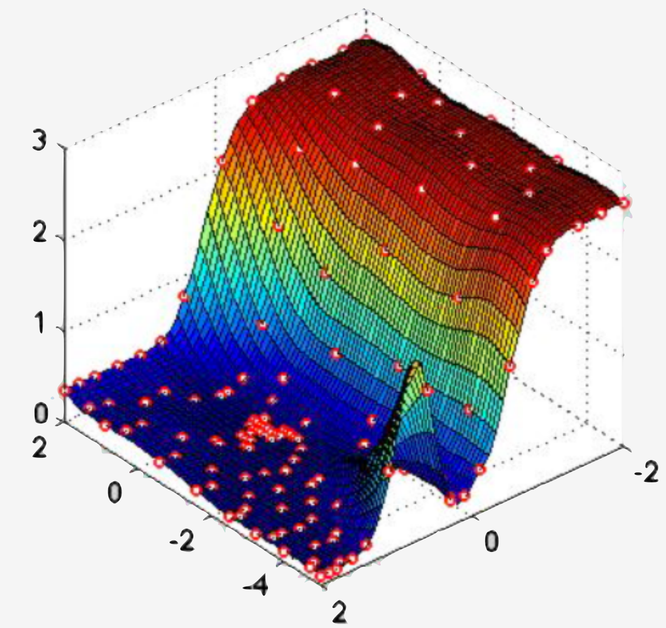
Problem: Releasing hyperparameters from grid search compromises privacy

[Chaudhuri & Vinterbo, 2013]

Solution: Design Diff. Private grid search


Grid search
end

Bayesian Optimization



Can we make Bayesian Optimization private?

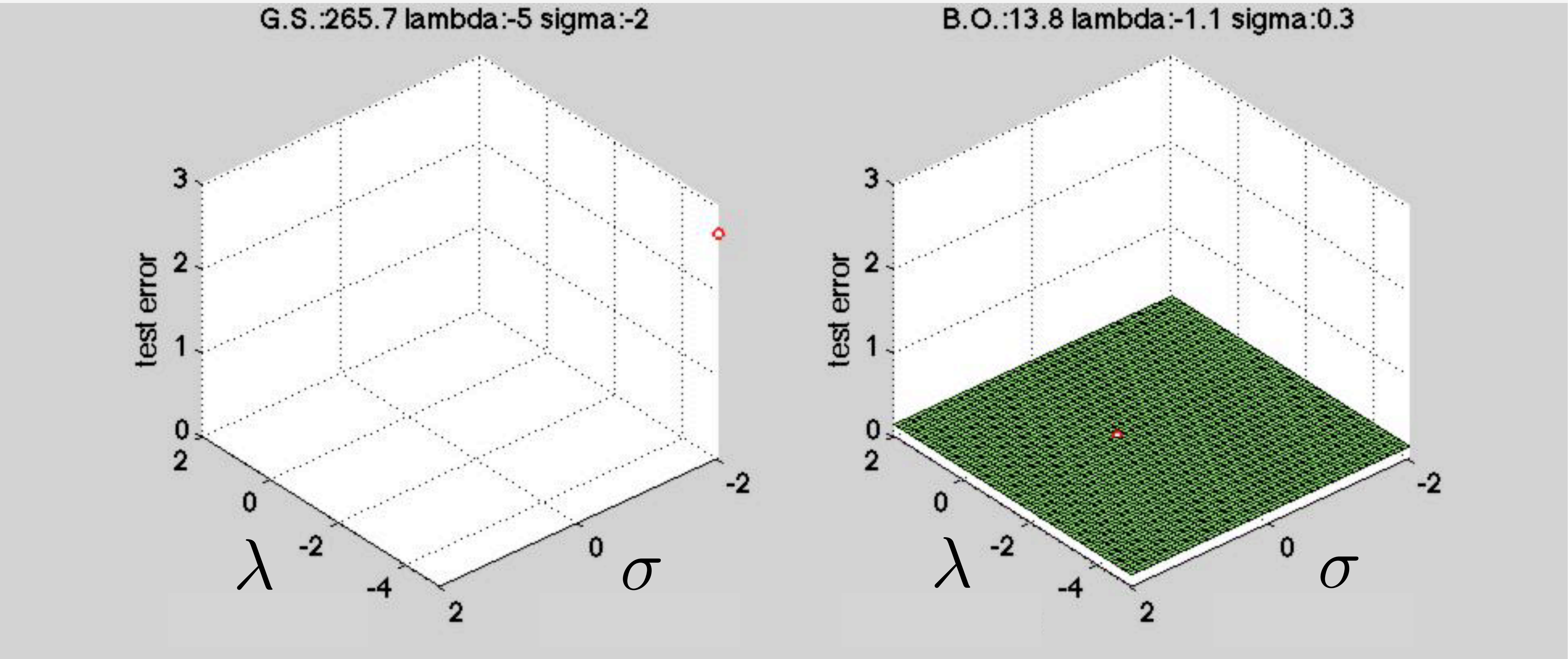
Bayesian Optimization

<u>A general hyperparameter tuning method</u>

[Hutter et al. 2011; Bergstra & Bengio, 2012; Snoek et al. 2012; Gardner et al., 2014]

e.g., RBF Kernel SVM has hyperparameters: (λ, σ)

regularization trade-off (λ) — kernel width (σ)



Bayesian Optimization

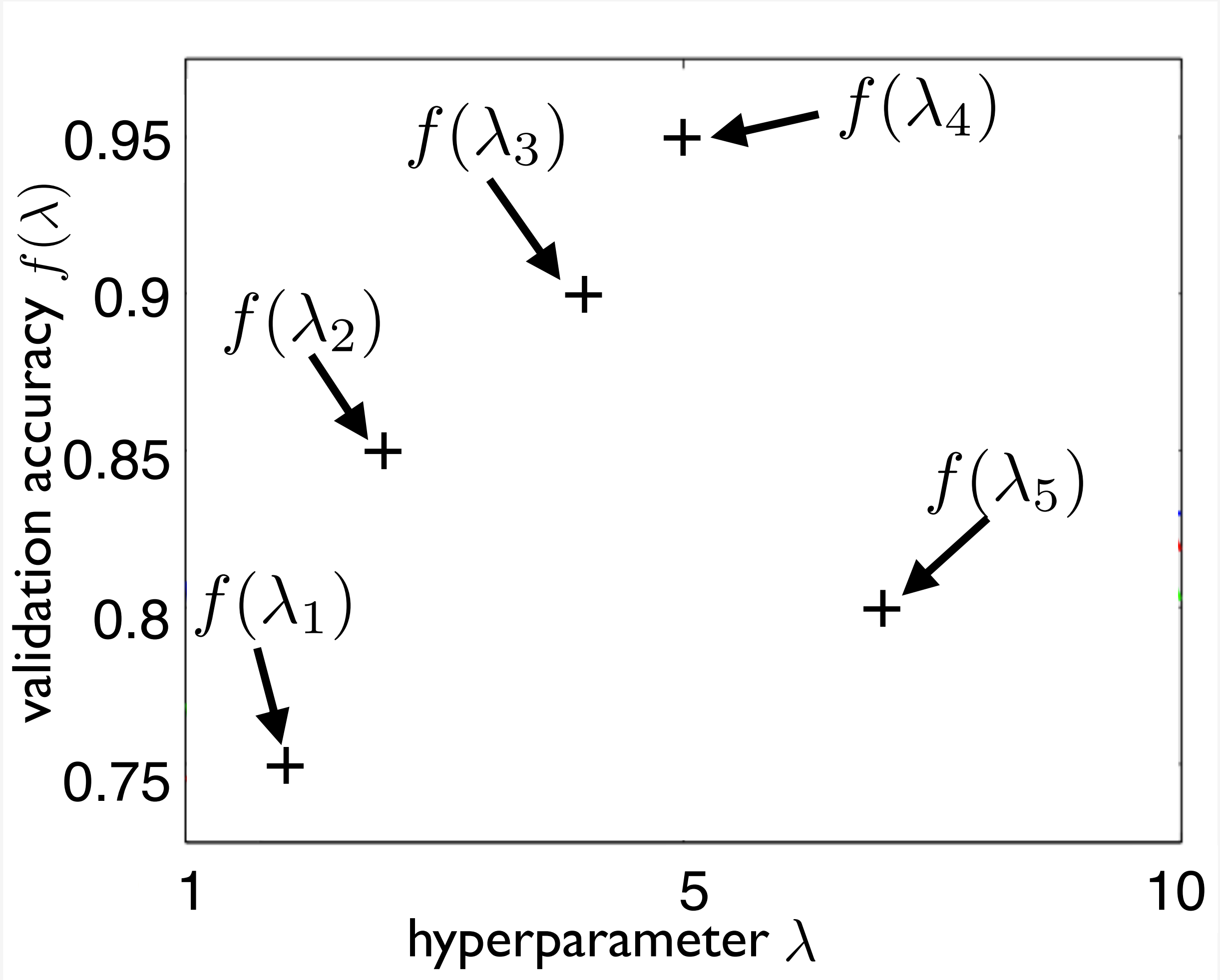


figure credit: [Rasmussen & Williams, 2006]

Bayesian Optimization

$$h(\hat{\lambda}) \sim \mathcal{N}(\mu(\hat{\lambda}), \sigma^2(\hat{\lambda}))$$

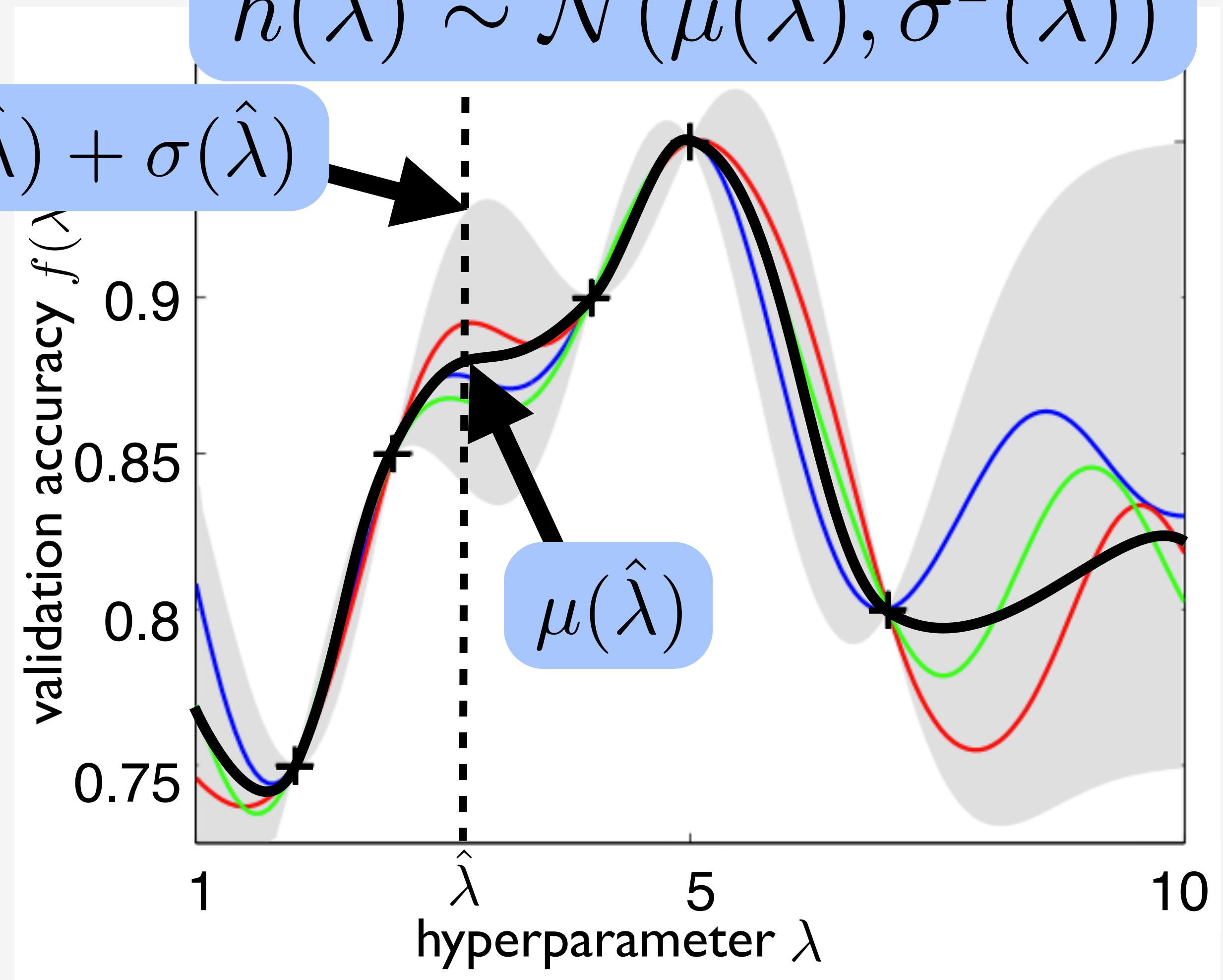


figure credit: [Rasmussen & Williams, 2006]

fit a Gaussian Process

$$h \sim \mathcal{GP}(0, k(\boldsymbol{\lambda}, \boldsymbol{\lambda}'))$$

Bayesian Optimization

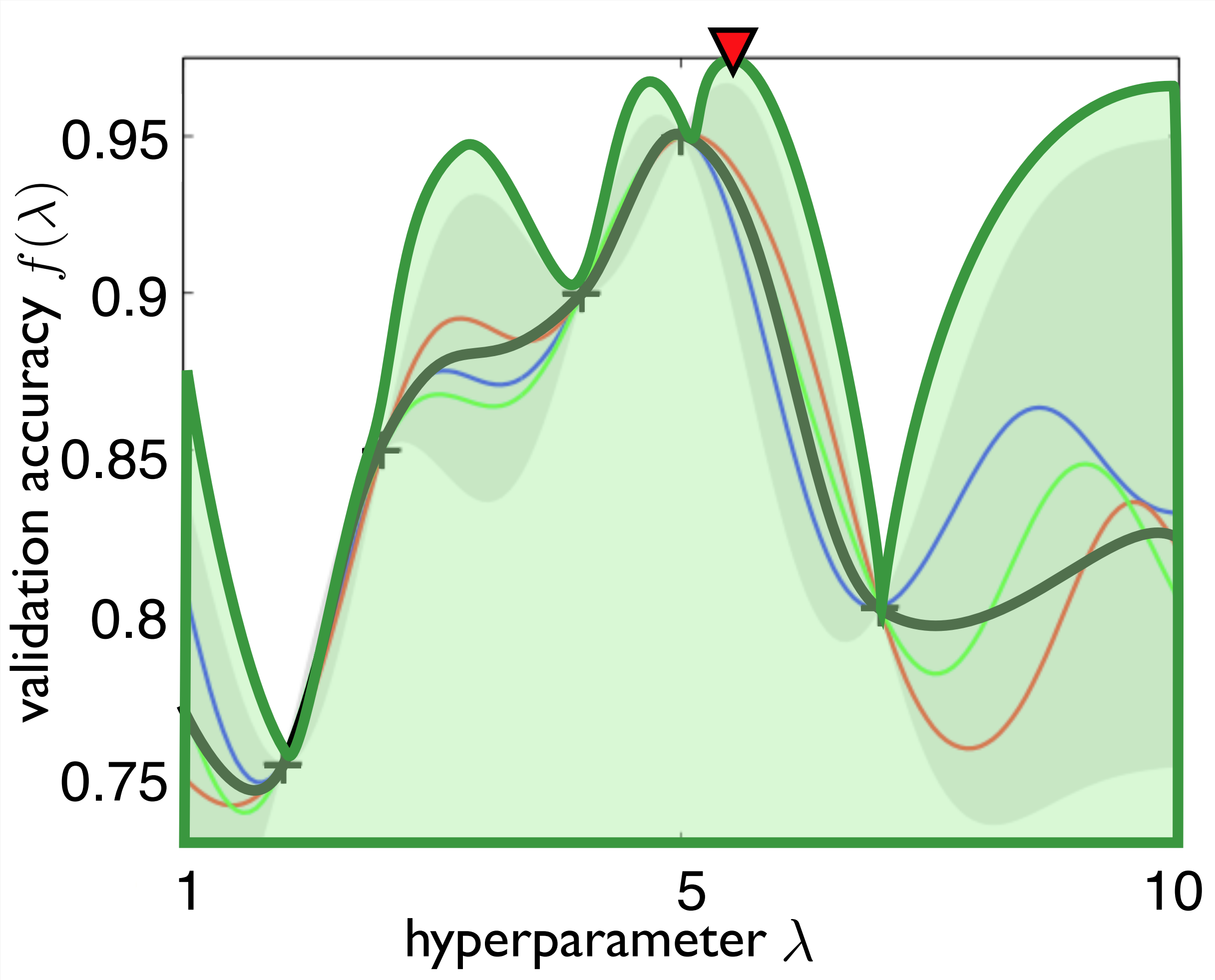


figure credit: [Rasmussen & Williams, 2006]

[Srinivas et al., 2010]

Maximize Upper Confidence Bound (GP-UCB)

$$\mu(\lambda) + \sqrt{\beta}\sigma(\lambda)$$

Differential Privacy

[Dwork et al., 2006]

A formalization of “privacy through randomness”

[Dwork et al., 2006]

Differential Privacy

A formalization of “privacy through randomness”

algorithm — $\mathcal{A}$

data — $\mathcal{V}$

age: 20	gender: M	SSN: 42...2
age: 91	gender: M	SSN: 21...0
age: 82	gender: F	SSN: 46...7
age: 39	gender: F	SSN: 91...3
⋮	⋮	⋮

output
e.g., validation accuracy

$$\mathcal{A}\left(\mathcal{V}\right) \longrightarrow f(\boldsymbol{\lambda})$$

Differential Privacy

[Dwork et al., 2006]

A formalization of "privacy through randomness"

algorithm

data $\mathcal{V}'$

output

e.g., validation accuracy

$\mathcal{A}$

age: 20	gender: M	SSN: 42...2
age: 91	gender: M	SSN: 21...0
age: 56	gender: F	SSN: 72...4
age: 39	gender: F	SSN: 91...3
⋮	⋮	⋮

$$\longrightarrow f(\lambda')$$

[Dwork et al., 2006]

Differential Privacy

A formalization of “privacy through randomness”

algorithm — data — output

e.g., validation accuracy

$\mathcal{V}'$

$\mathcal{A}\Big($

age: 20	gender: M	SSN: 42...2
age: 91	gender: M	SSN: 21...0
age: 56	gender: F	SSN: 72...4
age: 39	gender: F	SSN: 91...3
⋮	⋮	⋮

$\Big) \longrightarrow f(\boldsymbol{\lambda}')$

informally:

$$f(\boldsymbol{\lambda}) \approx f(\boldsymbol{\lambda}')$$

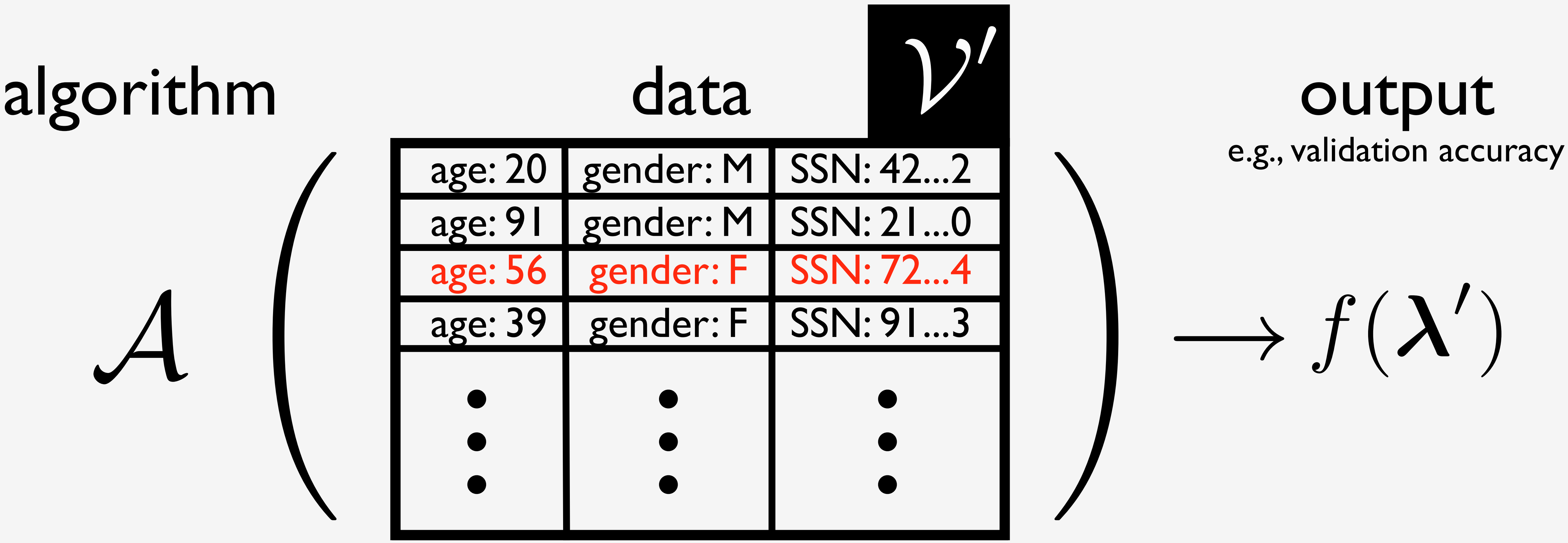
[in certain settings]

$$\boldsymbol{\lambda} \approx \boldsymbol{\lambda}'$$

Differential Privacy

[Dwork et al., 2006]

A formalization of “privacy through randomness”



Definition 1. *A randomized algorithm $\mathcal{A}$ is (ϵ, δ)-**differentially private** for $\epsilon, \delta \geq 0$ if for all $f(\lambda) \in \text{Range}(\mathcal{A})$ and for all neighboring datasets $\mathcal{V}, \mathcal{V}'$ (i.e., such that $\mathcal{V}$ and $\mathcal{V}'$ differ in the value of one record) we have that*

$$\Pr\left[\mathcal{A}(\mathcal{V}) = f(\lambda)\right] \leq e^{\epsilon} \Pr\left[\mathcal{A}(\mathcal{V}') = f(\lambda)\right] + \delta.$$

Private Mechanisms

Definition 2. *(Laplace mechanism) The* ***global sensitivity*** *of an algorithm* $\mathcal{A}$ *over all neighboring datasets* $\mathcal{V}, \mathcal{V}'$ *(*$\mathcal{V}, \mathcal{V}'$ *differ by the value of one record) is*

$$\Delta_{\mathcal{A}} \triangleq \max_{\mathcal{V}, \mathcal{V}' \subseteq \mathcal{X}} \|\mathcal{A}(\mathcal{V}) - \mathcal{A}(\mathcal{V}')\|_1 .$$

(Exponential mechanism) The ***global sensitivity*** *of a function* $q : \mathcal{X} \times \Lambda \rightarrow \mathbb{R}$ *over all neighboring datasets* $\mathcal{V}, \mathcal{V}'$ *is*

$$\Delta_q \triangleq \max_{\substack{\mathcal{V}, \mathcal{V}' \subseteq \mathcal{X} \\ \lambda \in \Lambda}} \|q(\mathcal{V}, \lambda) - q(\mathcal{V}', \lambda)\|_1 .$$

Laplace Mechanism

[Dwork et al., 2006]

1. Draw $\omega \sim \mathrm{Laplace}(0, \Delta_{\mathcal{A}}/\epsilon)$
2. Release $\mathcal{A}(\mathcal{V}) + \omega$

The Laplace Mechanism is $(\epsilon, 0)$-differentially private!

Exponential Mechanism

[McSherry & Talwar, 2007]

1. Draw $\tilde{\lambda} \sim \frac{1}{Z} \exp(\epsilon q(\mathcal{V}, \lambda)/(2\Delta_q))$
2. Release $\tilde{\lambda}$

The Exponential Mechanism is $(\epsilon, 0)$-differentially private!

GP Assumption

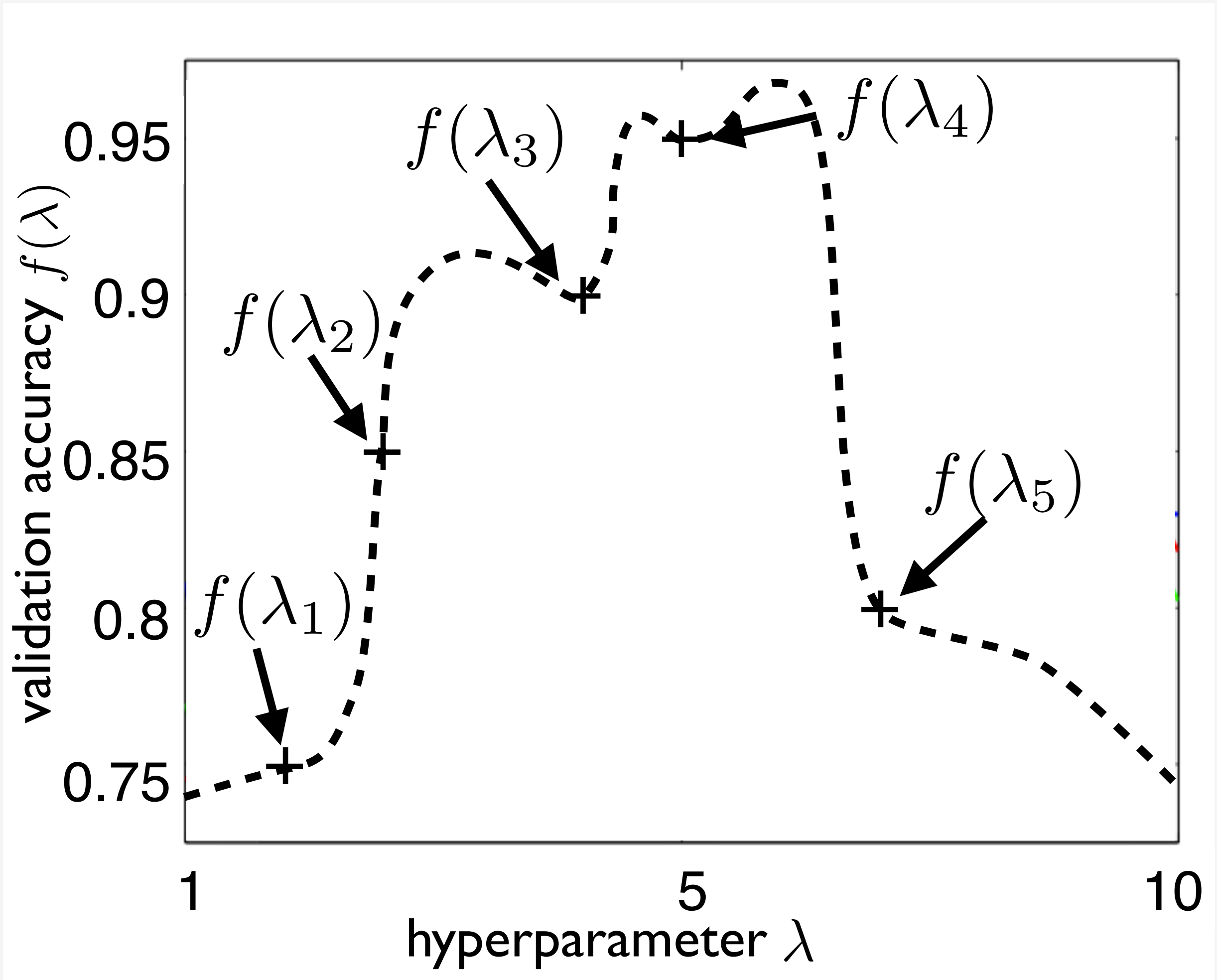


figure credit: [Rasmussen & Williams, 2006]

suppose...

$$f \sim \mathcal{GP}(0, k(\boldsymbol{\lambda}, \boldsymbol{\lambda}'))$$

GP Assumption

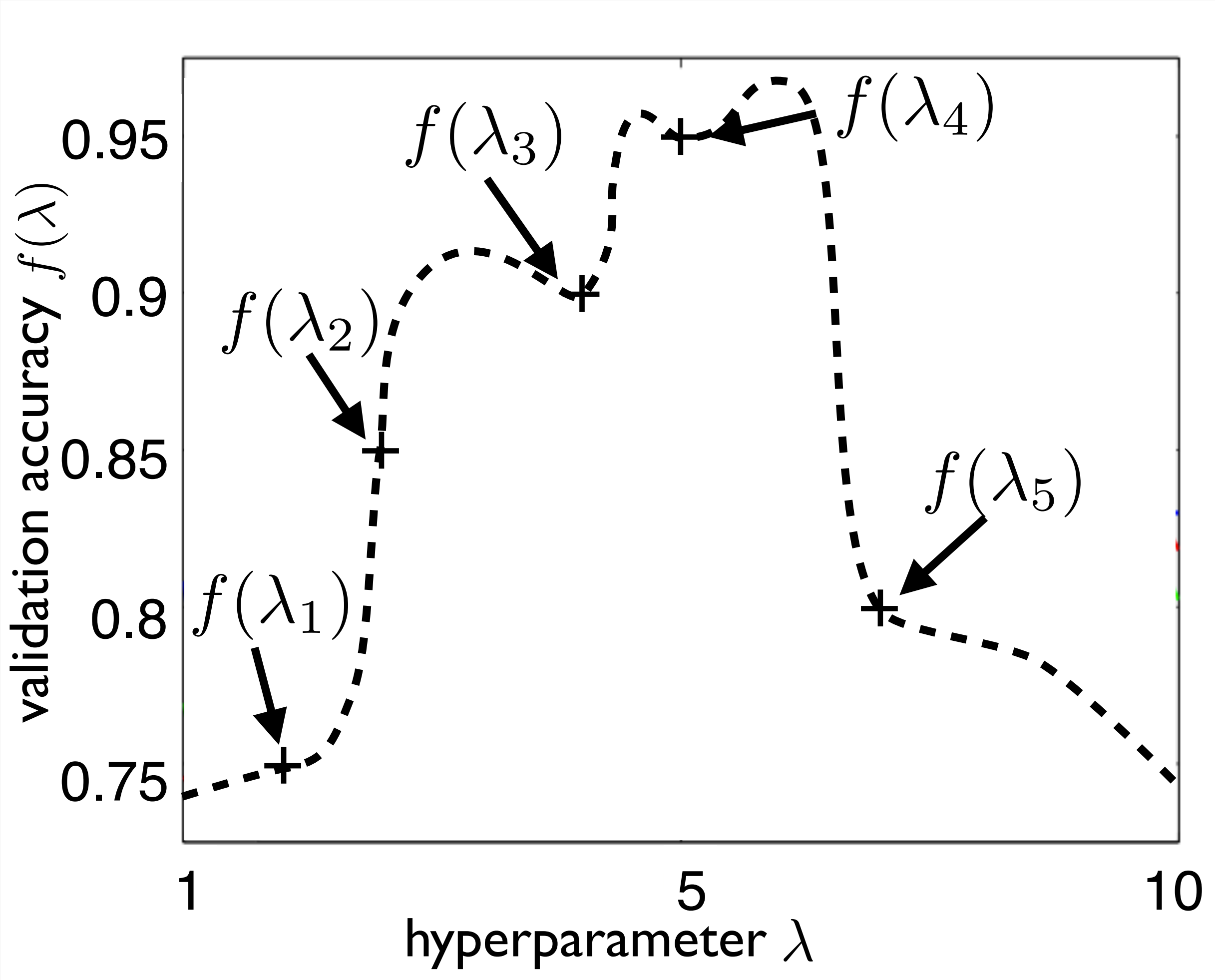


figure credit: [Rasmussen & Williams, 2006]

suppose... $f \sim \mathcal{GP}(0, k(\boldsymbol{\lambda}, \boldsymbol{\lambda}'))$

[Srinivas et al., 2010]

bounded regret! $\frac{1}{T}\sum_{t=1}^{T} f(\lambda^*) - f(\lambda_t) \leq O\Big(\frac{1}{\sqrt{T}}\Big)$

Assumption for Privacy

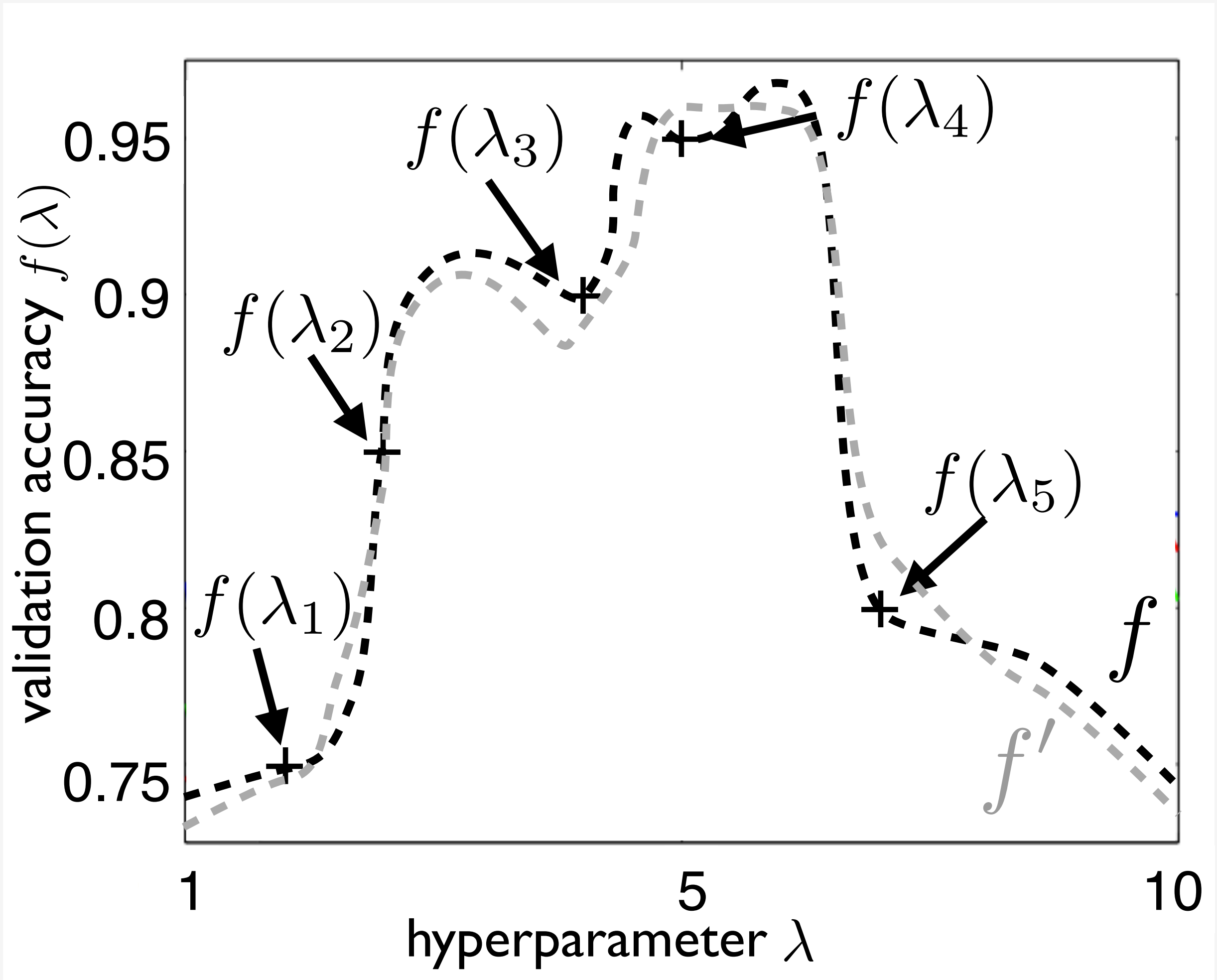


figure credit: [Rasmussen & Williams, 2006]

suppose... $[f, f'] \sim \mathcal{GP}(0, k_1(\mathcal{V}, \mathcal{V}') \otimes k_2(\boldsymbol{\lambda}, \boldsymbol{\lambda}'))$

dataset kernel

hyperparam. kernel

Assumption for Privacy

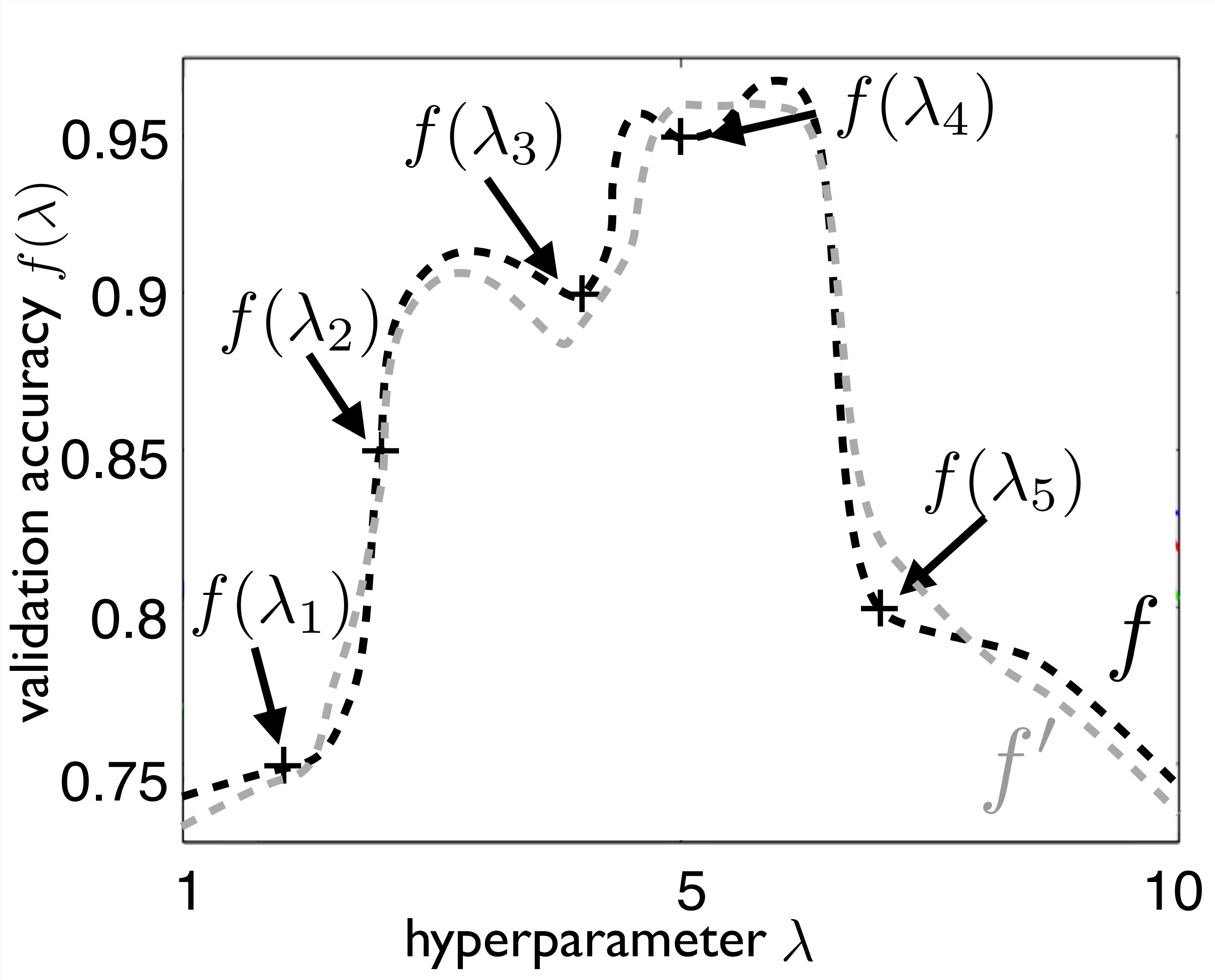


figure credit: [Rasmussen & Williams, 2006]

suppose... $[f, f'] \sim \mathcal{GP}(0, k_1(\mathcal{V}, \mathcal{V}') \otimes k_2(\boldsymbol{\lambda}, \boldsymbol{\lambda}'))$

Differential Privacy + Utility!

Theorems 1 & 2

Release: Private hyperparameter values

Setting: We observe noisy validation accuracies

Main idea: Run GP-UCB and use Exp. Mechanism

[McSherry & Talwar, 2007]

1. How to use Exp. Mechanism?

Theorem 1. *Given the GP assumption, for any two neighboring datasets $\mathcal{V}, \mathcal{V}'$ and for all $\lambda \in \Lambda$ with probability at least $1-\delta$ there is an upper bound on the global sensitivity of μ_T:*

$$|\mu'_T(\lambda) - \mu_T(\lambda)| \leq O\Big(\sqrt{\log(|\Lambda|(T+1)^2/\delta)} + \sqrt{(1-k_1(\mathcal{V},\mathcal{V}'))\log(|\Lambda|/\delta)}\Big)$$

Theorems 1 & 2

Release: Private hyperparameter values

Setting: We observe noisy validation accuracies

Main idea: Run GP-UCB and use Exp. Mechanism

[McSherry & Talwar, 2007]

1. How to use Exp. Mechanism?

Theorem 1. *Given the GP assumption, for any two neighboring datasets $\mathcal{V}, \mathcal{V}'$ and for all $\lambda \in \Lambda$ with probability at least $1 - \delta$ there is an upper bound on the global sensitivity of μ_T:*

$$|\mu'_T(\lambda) - \mu_T(\lambda)| \leq O\left(\sqrt{\log(|\Lambda|(T+1)^2/\delta)} + \sqrt{(1 - k_1(\mathcal{V}, \mathcal{V}'))\log(|\Lambda|/\delta)}\right)$$

Theorems 1 & 2

Release: Private hyperparameter values

Setting: We observe noisy validation accuracies

Main idea: Run GP-UCB and use Exp. Mechanism [McSherry & Talwar, 2007]

1. How to use Exp. Mechanism?

Theorem 1. *Given the GP assumption, for any two neighboring datasets $\mathcal{V}, \mathcal{V}'$ and for all $\lambda \in \Lambda$ with probability at least $1-\delta$ there is an upper bound on the global sensitivity of μ_T:*

$$|\mu'_T(\lambda) - \mu_T(\lambda)| \leq O\left(\sqrt{\log(|\Lambda|(T+1)^2/\delta)} + \sqrt{(1-k_1(\mathcal{V},\mathcal{V}'))\log(|\Lambda|/\delta)}\right)$$

Theorems 1 & 2

Release: Private hyperparameter values

Setting: We observe noisy validation accuracies

Main idea: Run GP-UCB and use Exp. Mechanism
[McSherry & Talwar, 2007]

1. How to use Exp. Mechanism?

Theorem 1. *Given the GP assumption, for any two neighboring datasets $\mathcal{V}, \mathcal{V}'$ and for all $\lambda \in \Lambda$ with probability at least $1-\delta$ there is an upper bound on the global sensitivity of μ_T:*

B

$$|\mu'_T(\lambda) - \mu_T(\lambda)| \leq O\left(\sqrt{\log(|\Lambda|(T+1)^2/\delta)} + \sqrt{(1-k_1(\mathcal{V},\mathcal{V}'))\log(|\Lambda|/\delta)}\right)$$

2. How good are the noisy hyperparameters?

Theorem 2. *(McSherry & Talwar, 2007) The exponential mechanism selects $\tilde{\lambda}$ that has value $\mu_T(\tilde{\lambda})$ that is close to the maximum $\max_{\lambda\in\Lambda} \mu_T(\lambda)$, w.p. $\geq 1-(\delta+e^{-a})$*

$$\max_{\lambda\in\Lambda} \mu_T(\lambda) - \mu_T(\tilde{\lambda}) \leq O\left(\frac{B}{\epsilon}(\log|\Lambda| + a)\right)$$

Theorems 3 & 4

Release: Private validation accuracies

Setting: We observe noisy validation accuracies

Main idea: Run GP-UCB and use Lap. Mechanism [Dwork et al., 2006]

1. How much noise to add?

Theorem 3. *Given the GP assumption, and neighboring $\mathcal{V}, \mathcal{V}'$, we have the following global sensitivity bound for the maximum v, w.p. $\geq 1-\delta$*

$$|\max_{t \leq T} v'_t - \max_{t \leq T} v_t| \leq O\Big(\frac{1}{\sqrt{T}} + \sqrt{(1 - k_1(\mathcal{V}, \mathcal{V}')) \log(|\Lambda|/\delta)} + \sqrt{\log(1/\delta)}\Big).$$

where $k_2(\lambda, \lambda')$ is the squared exponential kernel.

Theorems 3 & 4

Release: Private validation accuracies

Setting: We observe noisy validation accuracies

Main idea: Run GP-UCB and use Lap. Mechanism [Dwork et al., 2006]

1. How much noise to add?

Theorem 3. *Given the GP assumption, and neighboring $\mathcal{V}, \mathcal{V}'$, we have the following global sensitivity bound for the maximum v, w.p. $\geq 1-\delta$*

$$|\max_{t \leq T} v'_t - \max_{t \leq T} v_t| \leq O\Big(\frac{1}{\sqrt{T}} + \sqrt{(1 - k_1(\mathcal{V}, \mathcal{V}')) \log(|\Lambda|/\delta)} + \sqrt{\log(1/\delta)}\Big).$$

where $k_2(\lambda, \lambda')$ is the squared exponential kernel.

B

2. How good is the noisy error?

Theorem 4. *Given the GP assumption we have, with probability at least $1-(\delta + e^{-a})$*

$$|\tilde{v} - f(\lambda^*)| \leq O\Big(\sqrt{\log(1/\delta)} + \tfrac{a+\epsilon}{\epsilon\sqrt{T}} + \tfrac{aB}{\epsilon}\Big).$$

Our Results

1. If $f(\boldsymbol{\lambda})$ satisfies Gaussian Process smoothness assumptions then,

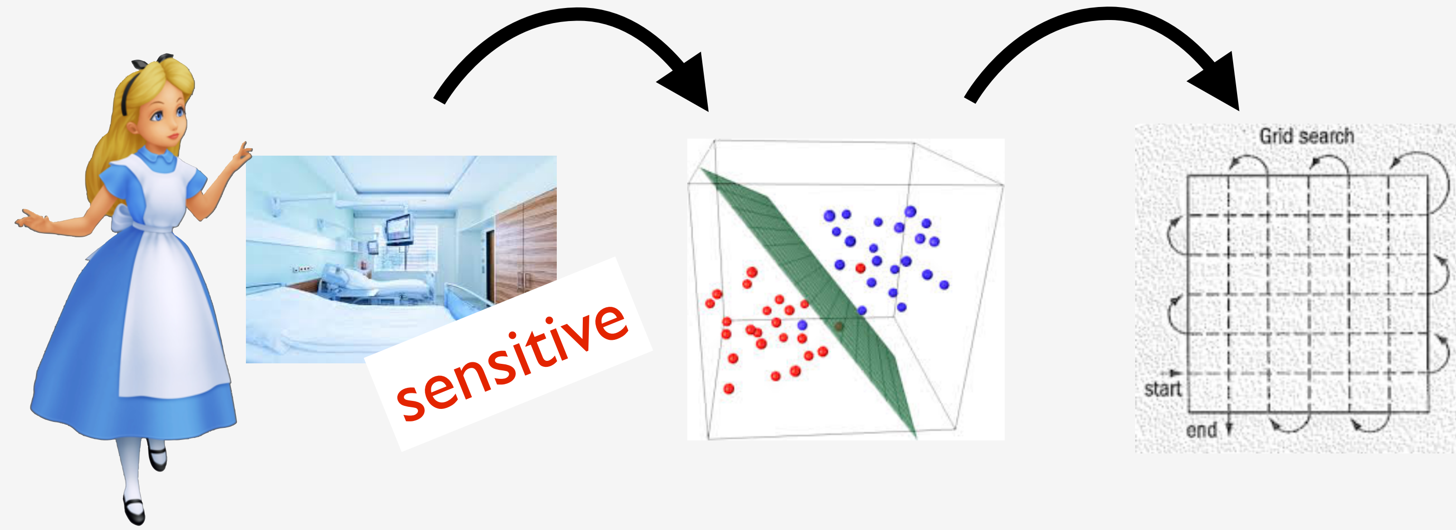
	private $f(\boldsymbol{\lambda})$	private $\boldsymbol{\lambda}$
exact observation [de Freitas et al., 2012]	✓	✗
with observation noise [Srinivas et al., 2010]	✓	✓

2. If $f(\boldsymbol{\lambda})$ satisfies Lipschitz smoothness and convexity assumptions then,

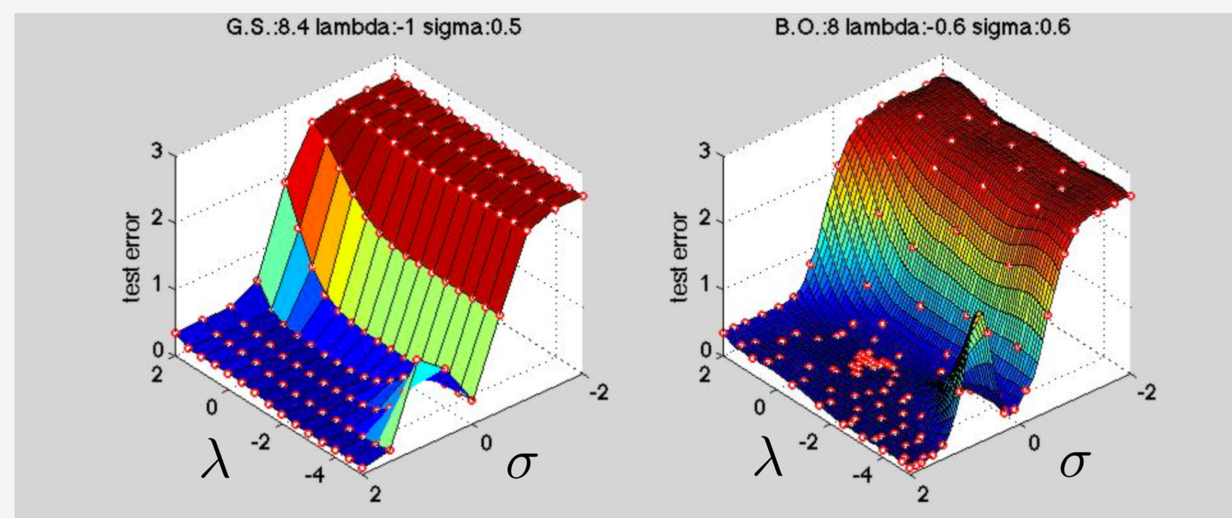
✓ private $f(\boldsymbol{\lambda})$ (exact observation) using any BO procedure!

Take Home Points

1. Releasing sensitive validation grid search results can compromise privacy
[Chaudhuri & Vinterbo, 2013]



2. Bayesian Optimization is the state-of-the-art for hyperparameter tuning



3. We present initial results for private Bayesian optimization

	private $f(\lambda)$	private λ
exact observation [de Freitas et al., 2012]	✓	✗
with observation noise [Srinivas et al., 2010]	✓	✓

Thank you. Questions?