Bayesian Fitting

Probabilistic Morphable Models

Summer School, June 2017

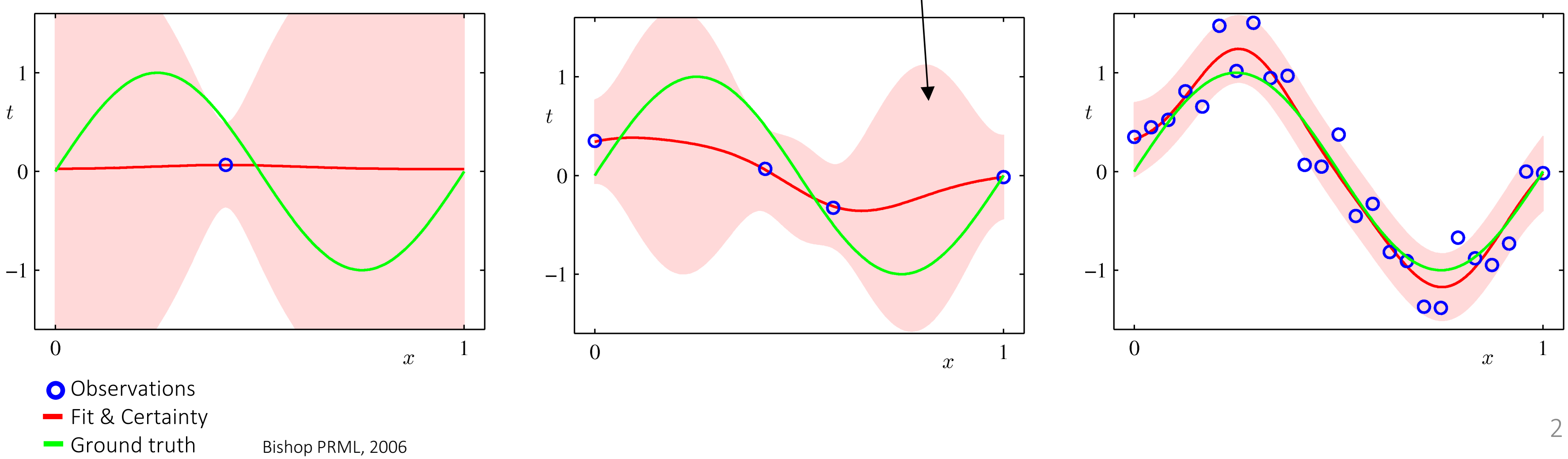
Sandro Schönborn

University of Basel

Uncertainty: Probability Distributions

- Probabilistic Models
- Uncertain Observation (noise, outlier, occlusion, ...)
- Fitting: Model explanation of observed data – probabilistic?

Tells us about the outcome's certainty!



Bishop PRML, 2006

Probability: An Example

- Dentist example: *Does the patient have a cavity?*

$$P(\text{cavity}) = 0.1$$

$$P(\text{cavity}|\text{toothache}) = 0.8$$

$$P(\text{cavity}|\text{toothache}, \text{gum problems}) = 0.4$$

Certainty

But the patient either has a cavity or does not

- *There is no 80% cavity!*
- *Having a cavity should not depend on whether the patient has a toothache or gum problems*

All these statements do not contradict each other, they summarize ***the dentist's knowledge*** about the patient

AIMA: Russell & Norvig, *Artificial Intelligence. A Modern Approach,* 3rd edition,

Uncertainty: Bayesian Probability

- How are probabilities to be interpreted?

  They are sometimes contradictory: Why does the distribution change when we have more data? Shouldn't there be a *real* distribution of $P(\theta)$?

- Bayesian probabilities rely on a *subjective* perspective:

  Probability is used to express our *current knowledge*. It can *change* when we learn or see more: With more data, we are more *certain* about our result.

  Subjectivity: There is no single, real underlying distribution. A probability distribution expresses our knowledge – It is different in different situations and for different observers since they have different knowledge.

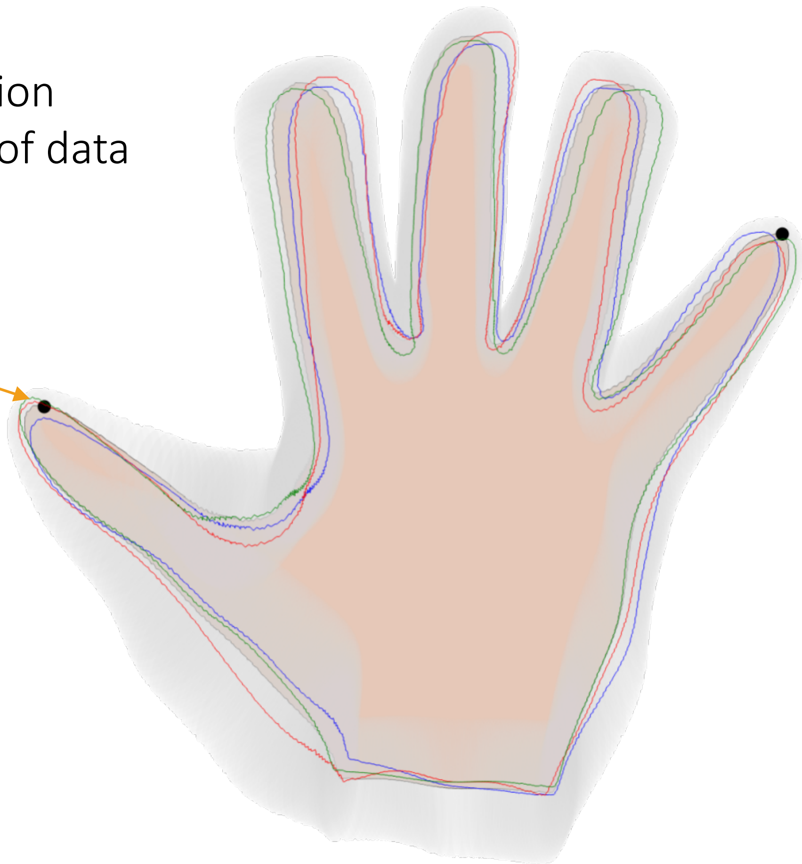
  - Not subjective in the sense that it is arbitrary!
    There are quantitative rules to follow mathematically

- Probability expresses an observers *certainty*, often called ***belief***

Towards Bayesian Inference

Posterior Models: Gaussian Process Regression
Probabilistic Fit: Probabilistic interpretation of data

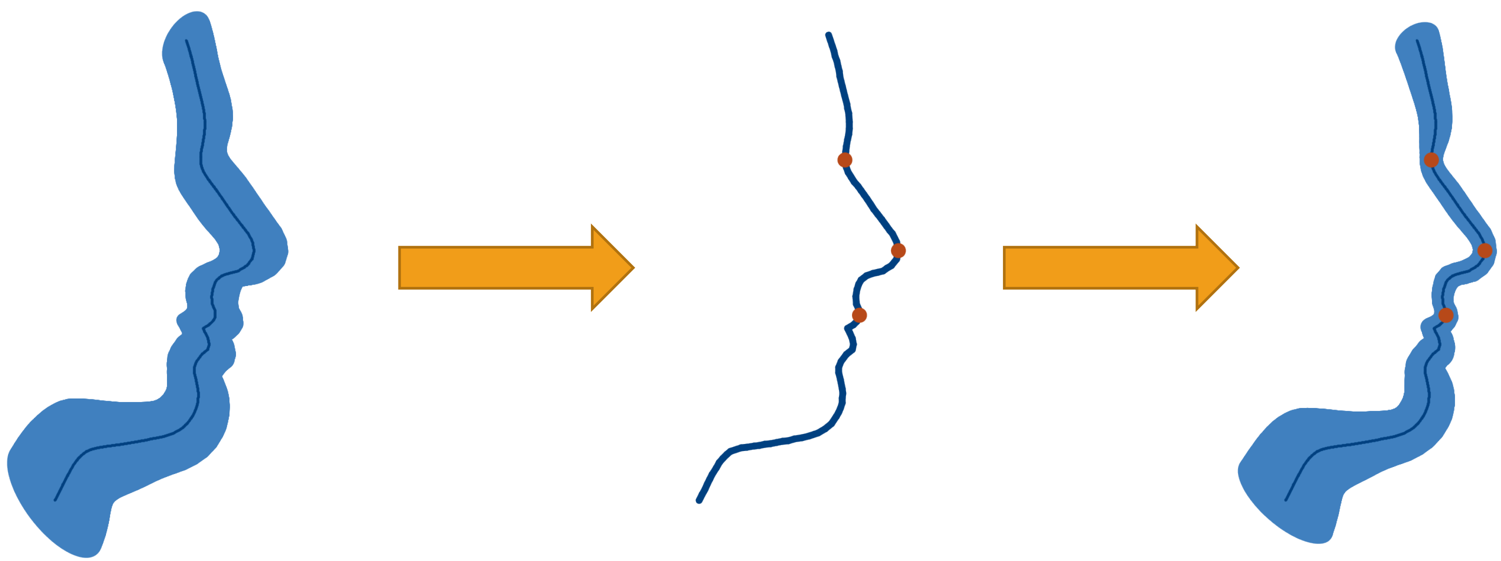
Observed Points



Update of prior to posterior model:
Bayesian Inference

Posterior Model

Belief Updates



Model
Face distribution

Observation
Concrete points
Possibly uncertain

Posterior
Face distribution
consistent with observation

Prior belief

More knowledge

Posterior belief

Consistency: Laws of probability calculus!

Joint Distribution

Probabilistic model: joint distribution of points

$$P(x_1, x_2)$$

Marginal

Distribution of certain points only

$$P(x_1) = \sum_{x_2} P(x_1, x_2)$$

Conditional

Distribution of points conditioned on *known* values of others

$$P(x_1|x_2) = \frac{P(x_1, x_2)}{P(x_2)}$$

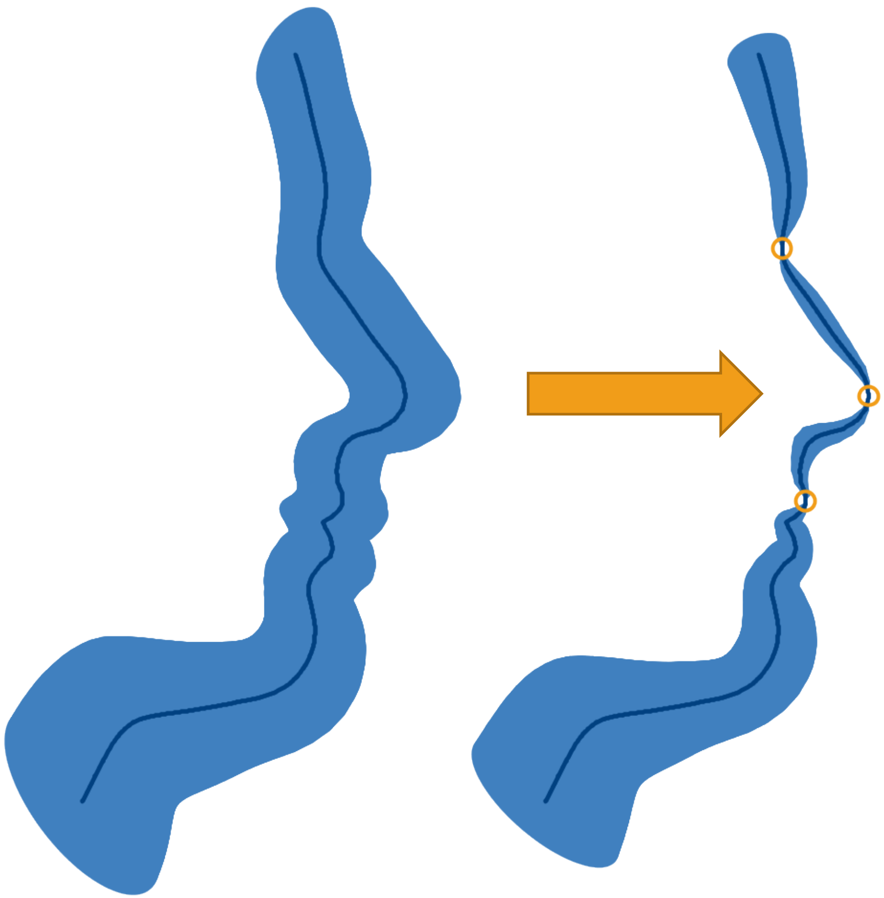
Both can be easily calculated for Gaussian models

Certain Observation

- Observations are *known* values
- Distribution of x_1 after *observing* $x_2, \ldots, x_N$:

$$P(x_1 | x_2 \ldots x_N) = \frac{P(x_1, x_2, \ldots, x_N)}{P(x_2, \ldots, x_N)}$$

- Conditional probability



Towards Bayesian Inference

- Update belief about x_1 by *observing* $x_2, \ldots, x_N$

$$P(x_1) \rightarrow P(x_1 | x_2 \ldots x_N)$$

- Factorize joint distribution

$$P(x_1, x_2, \ldots, x_N) = P(x_2, \ldots, x_N | x_1) P(x_1)$$

- Rewrite conditional distribution

$$P(x_1 | x_2 \ldots x_N) = \frac{P(x_1, x_2, \ldots, x_N)}{P(x_2, \ldots, x_N)} = \frac{P(x_2, \ldots, x_N | x_1) P(x_1)}{P(x_2, \ldots, x_N)}$$

- General: Query (Q) and Evidence (E)

$$P(Q|E) = \frac{P(Q, E)}{P(E)} = \frac{P(E|Q) P(Q)}{P(E)}$$

Uncertain Observation

- Observations with *uncertainty*

  Model needs to describe how observations are distributed

  with joint distribution $P(Q, E)$

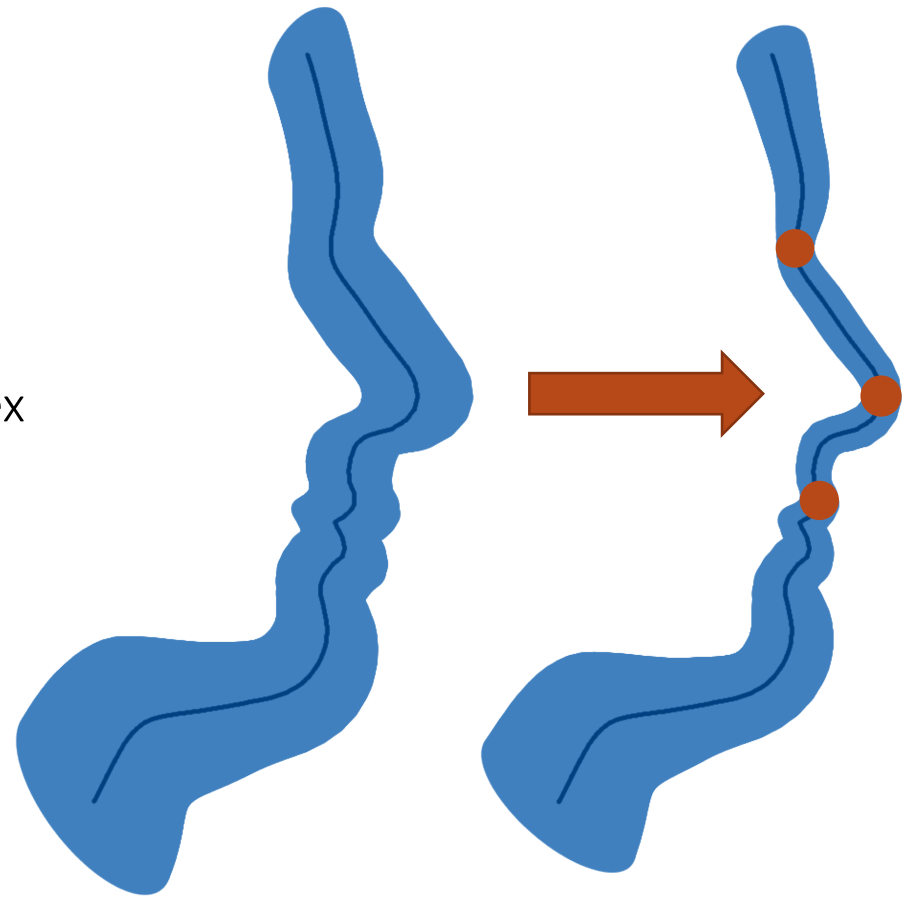
- Still conditional probability

  But joint distribution is more complex

- Joint distribution factorized

$$P(Q, E) = P(E|Q)P(Q)$$

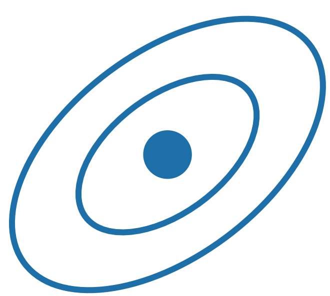
  - Likelihood $P(E|Q)$
  - Prior $P(Q)$



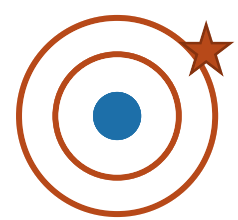
Likelihood

$$\underbrace{P(Q,E)}_{\text{Joint}} = \underbrace{P(E|Q)}_{\text{Likelihood}}\underbrace{P(Q)}_{\text{Prior}}$$

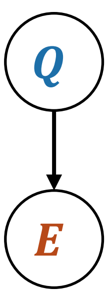
- *Likelihood x prior:* factorization is more flexible than full joint
  - Prior: distribution of core model *without observation*
  - Likelihood: describes how observations are distributed
- Common example: Gaussian distributed points



$P(Q)$



$P(E|Q)$



Bayesian Inference

- Conditional/Bayes rule: method to update *beliefs*

$$\underbrace{P(Q|E)}^{\text{Posterior}} = \frac{\overbrace{P(E|Q)}^{\text{Likelihood}}\overbrace{P(Q)}^{\text{Prior}}}{\underbrace{P(E)}_{\text{Marginal Likelihood}}}$$

- Each observation updates our belief (changes knowledge!)

$$P(Q) \rightarrow P(Q|E) \rightarrow P(Q|E,F) \rightarrow P(Q|E,F,G) \rightarrow \cdots$$

- Bayesian Inference: How beliefs *evolve* with observation
- Recursive: Posterior becomes prior of next inference step

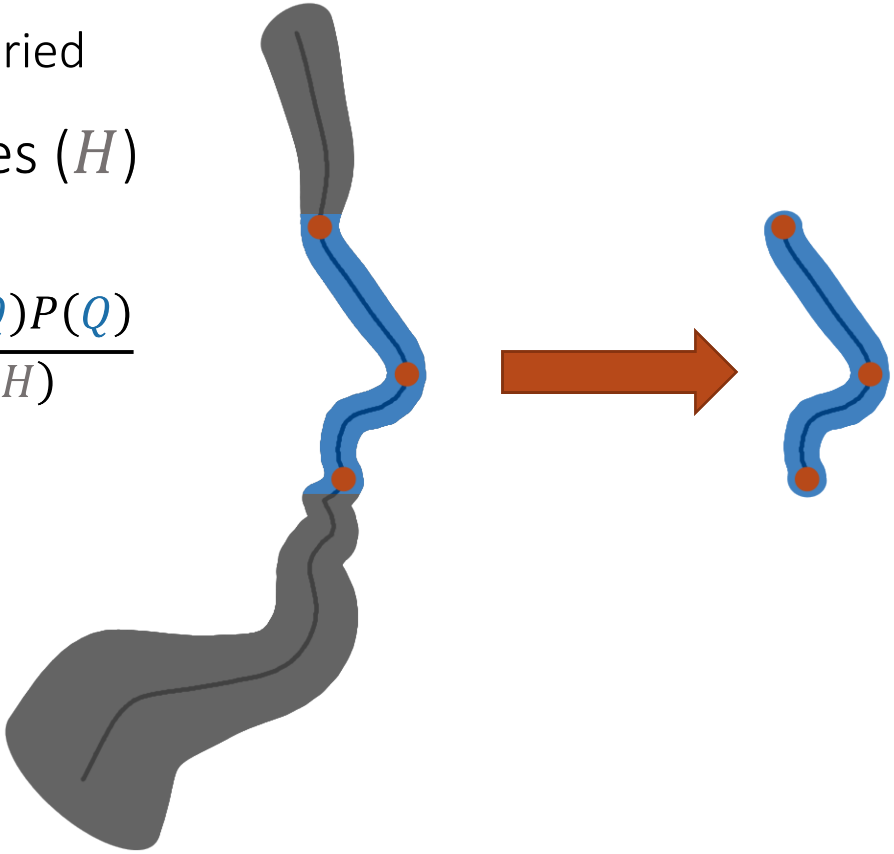
Marginalization

- Models contain irrelevant/hidden variables

  e.g. points on chin when nose is queried

- Marginalize over hidden variables (H)

$$P(Q|E) = \sum_H P(Q, H|E) = \sum_H \frac{P(E, H|Q)P(Q)}{P(E, H)}$$

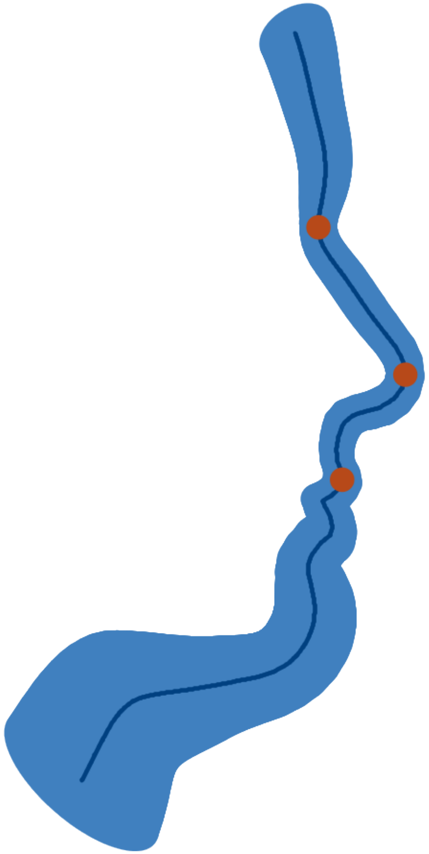


General Bayesian Inference

- Observation of *additional* variables
  - Common case, e.g. face rendering, landmark locations
  - Coupled to core model via likelihood factorization
- General Bayesian inference case:
  - Distribution of data D (formerly Evidence)
  - Parameters θ (formerly Query)

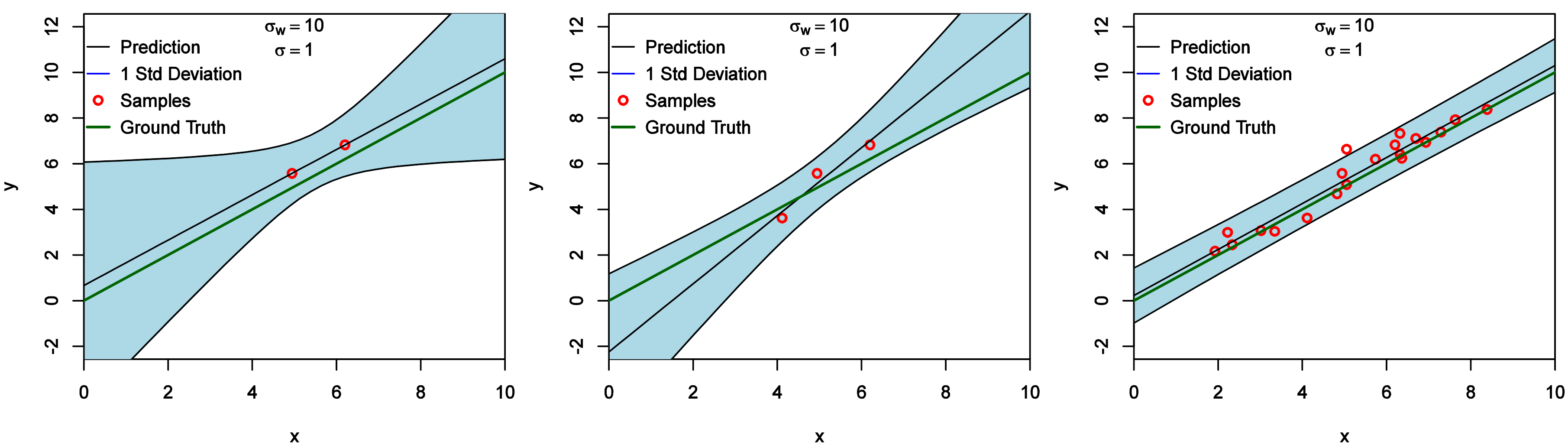
$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

$$P(\theta|D) \propto P(D|\theta)P(\theta)$$



Example: Bayesian Curve Fitting

- Curve Fitting: Data interpretation with a model
- Posterior distribution expresses certainty
  - in parameter space
  - in the predictive distribution



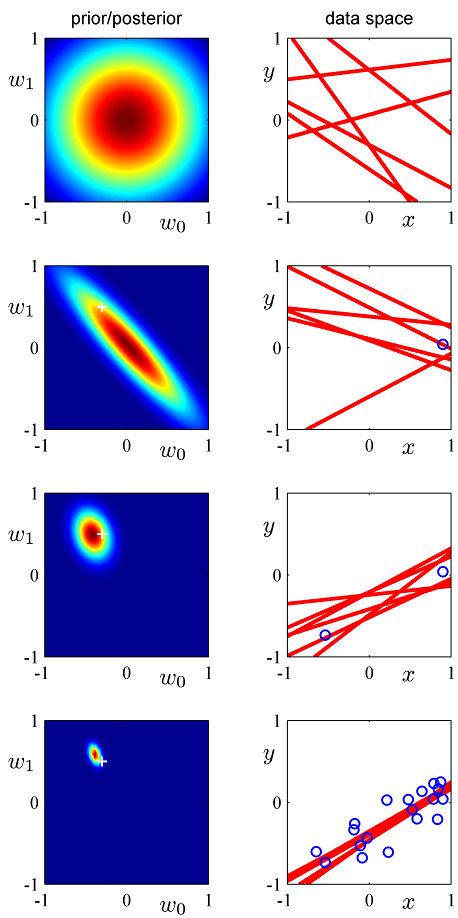
Posterior of Regression Parameters

No data

N=1

N=2

N=19



Bishop PRML, 2006

$$P(w)$$

$$\downarrow P(D_1|w)$$

$$P(w|D_1)$$

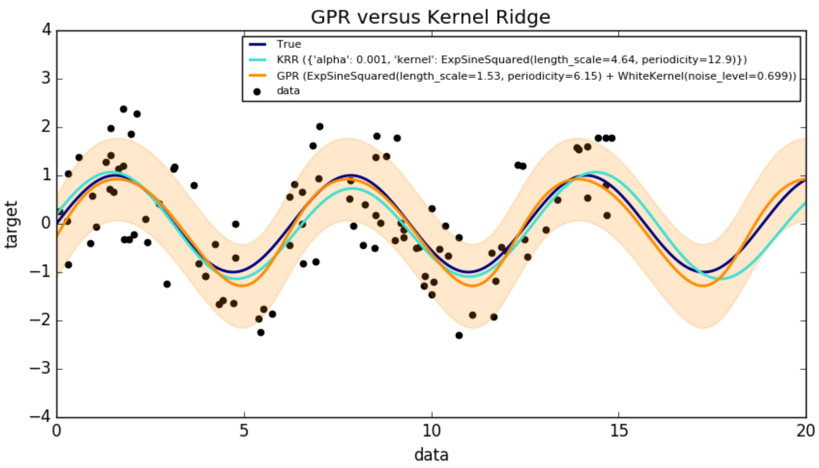
$$\downarrow P(D_2|w)$$

$$P(w|D_1, D_2)$$

$$\vdots$$

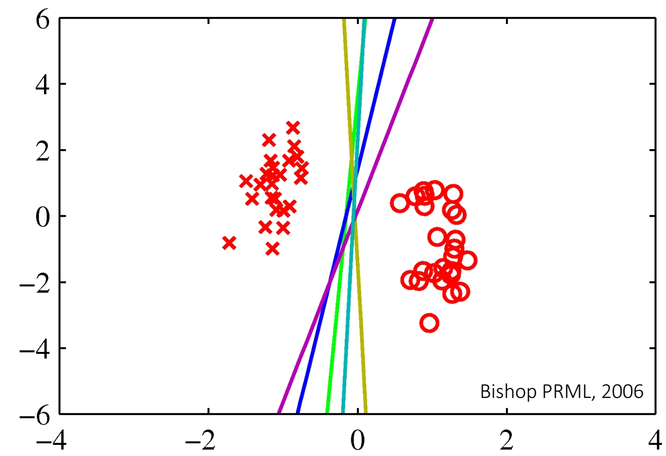
$$P(w|D_1, D_2, \dots)$$

More Bayesian Inference Examples



Non-Linear Curve Fitting

e.g. *Gaussian Process Regression*



Classification

e.g. *Bayes classifier*

Summary: Bayesian Inference

- *Belief*: formal expression of an *observer's knowledge*
  - Subjective state of knowledge about the world
- Beliefs are expressed as *probability* distributions
  - Formally not arbitrary: Consistency requires laws of probability
- *Observations* change knowledge and thus beliefs
- Bayesian inference formally updates *prior beliefs* to *posteriors*
  - Conditional Probability
  - Integration of observation via *likelihood x prior* factorization

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$