Computer vision: models, learning and inference

Chapter 6

Learning and Inference in Vision

Outline

- Computer vision models
  - Discriminative vs generative
- Worked example 1: Regression
- Worked example 2: Classification
- Which type should we choose?
- Applications

Computer Vision Inference

- Observe measured data, **x**
- Draw inferences from it about state of "world", **w**

Examples:

- Observe adjacent frames in video sequence
- Infer camera motion

- Observe image of face
- Infer identity

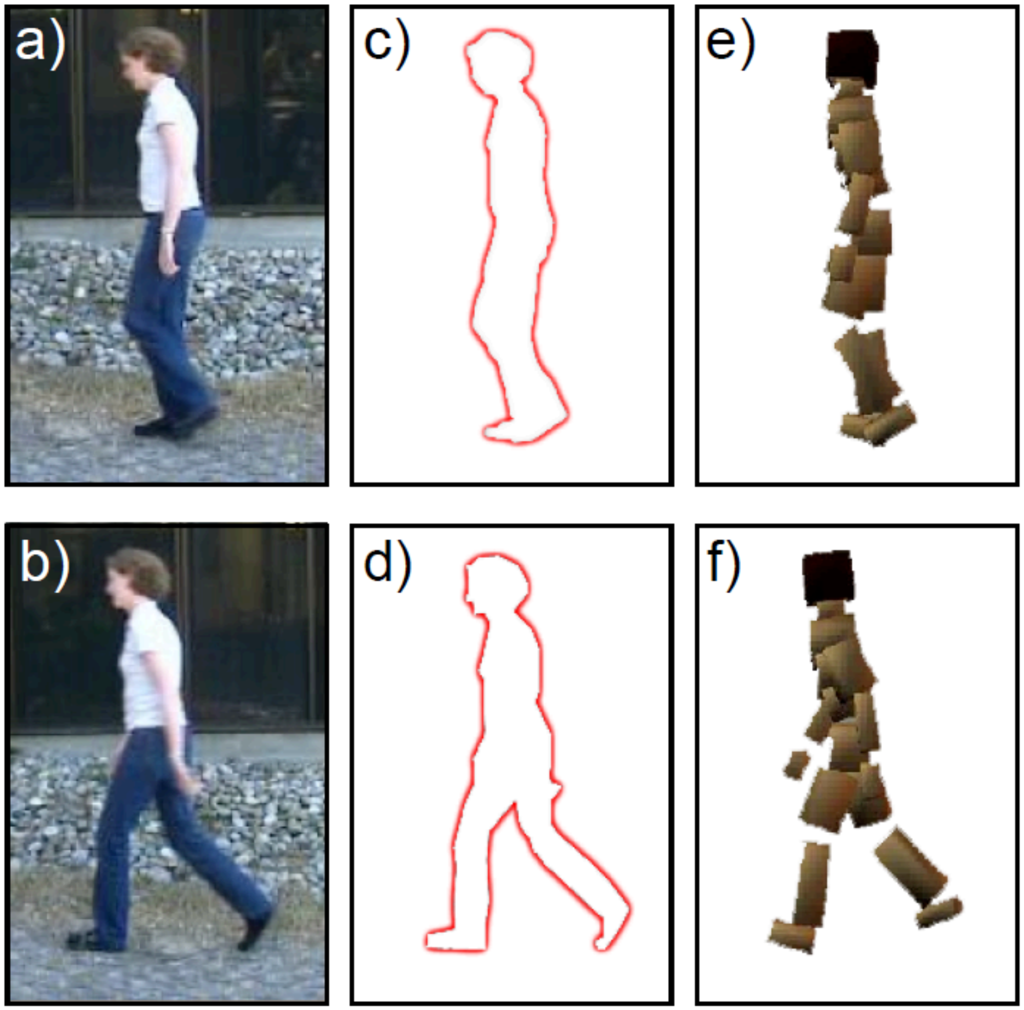
- Observe images from two displaced cameras
- Infer 3d structure of scene

Regression vs. Classification

- Observe measured data, **x**
- Draw inferences from it about world, **w**

When the world state **w** is continuous we'll call this regression

Regression Example
Pose from Image Silhouette



given measured silhouette
and a geometric body model

infer joint angles
(continuous quantities)

Regression Example
Pose from Image+Depth

Traditional RGB image

Image from new depth sensing camera

Body parts inferred by our recognition algorithm

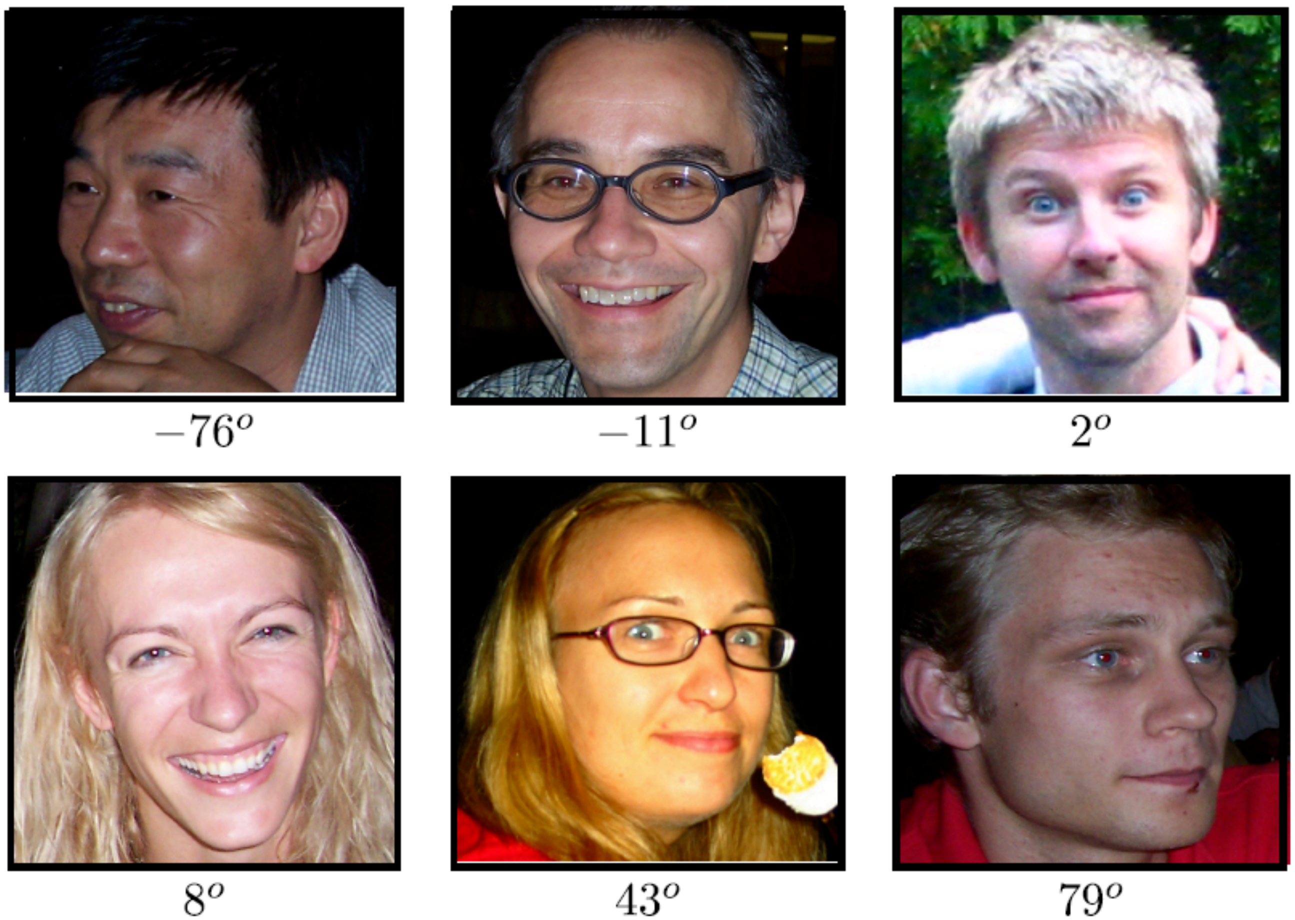
3D body part position proposals

Microsoft Kinect Sensor

given color image and registered depth image (and a geometric body model)

infer 3D joint positions

Regression Example
Head pose estimation



given measured image features

infer relative orientation angle of the head

Regression vs. Classification

- Observe measured data, **x**
- Draw inferences from it about world, **w**

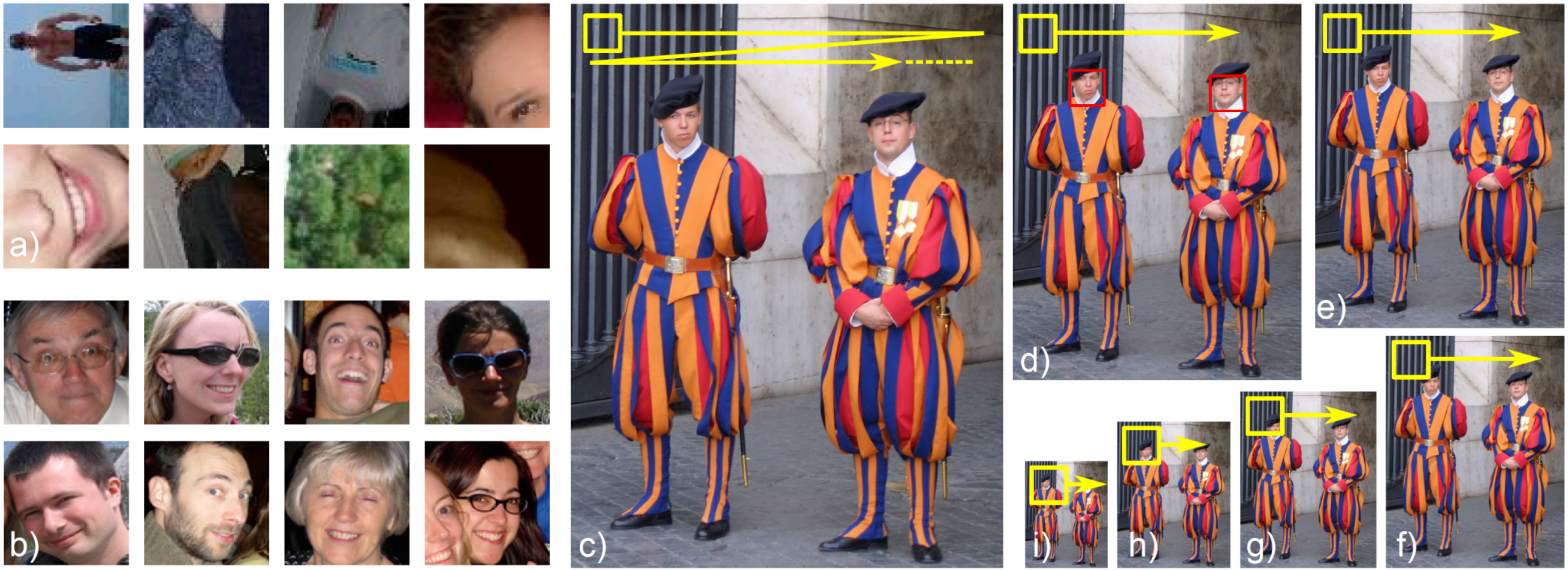
When the world state **w** is continuous we'll call this regression

When the world state **w** is discrete we call this classification

Classification Example

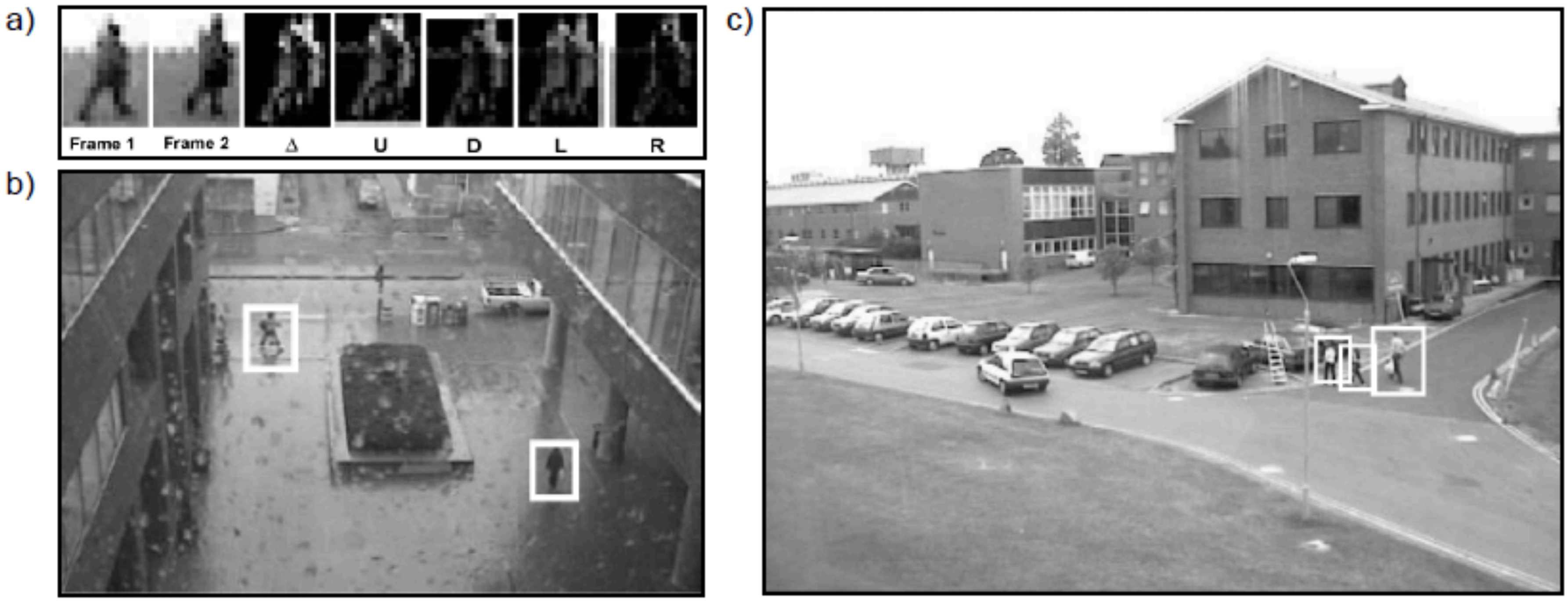
Face Detection

yes or no result: is there a face in the image?



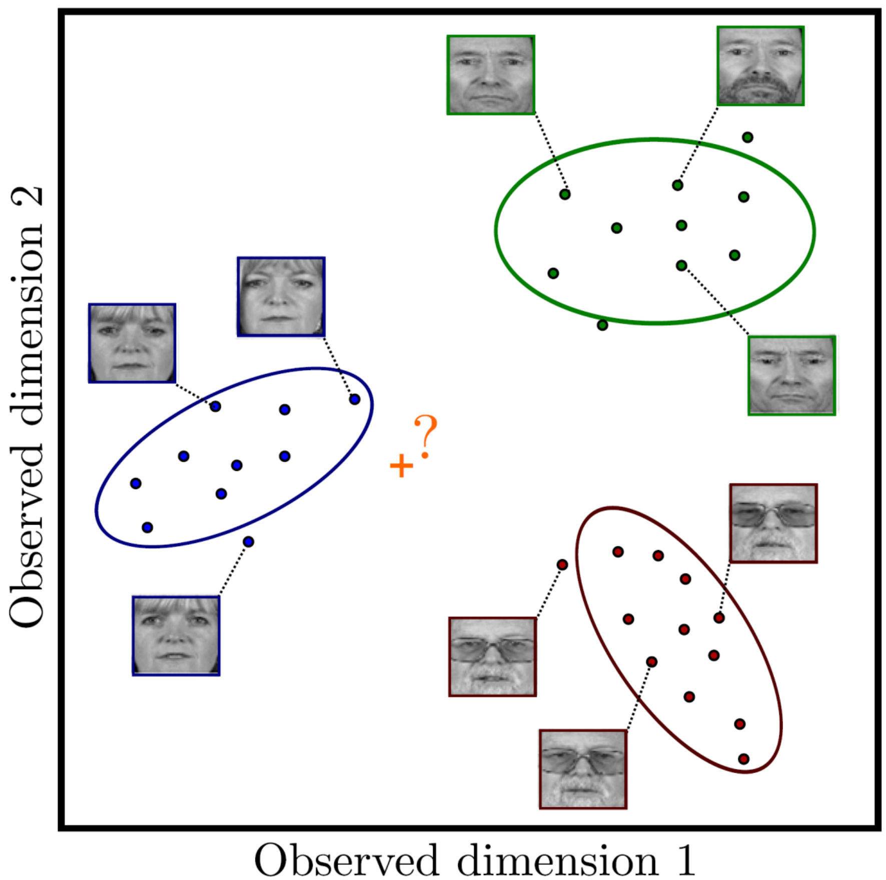
if you want to know where the face is, or count multiple faces, run the classifier over a sliding window of image patches

Classification Example:
Pedestrian Detection



same idea as face detection. HoG feature detector by Dalaal and Triggs is most common implementation.

Classification Example
Face Recognition



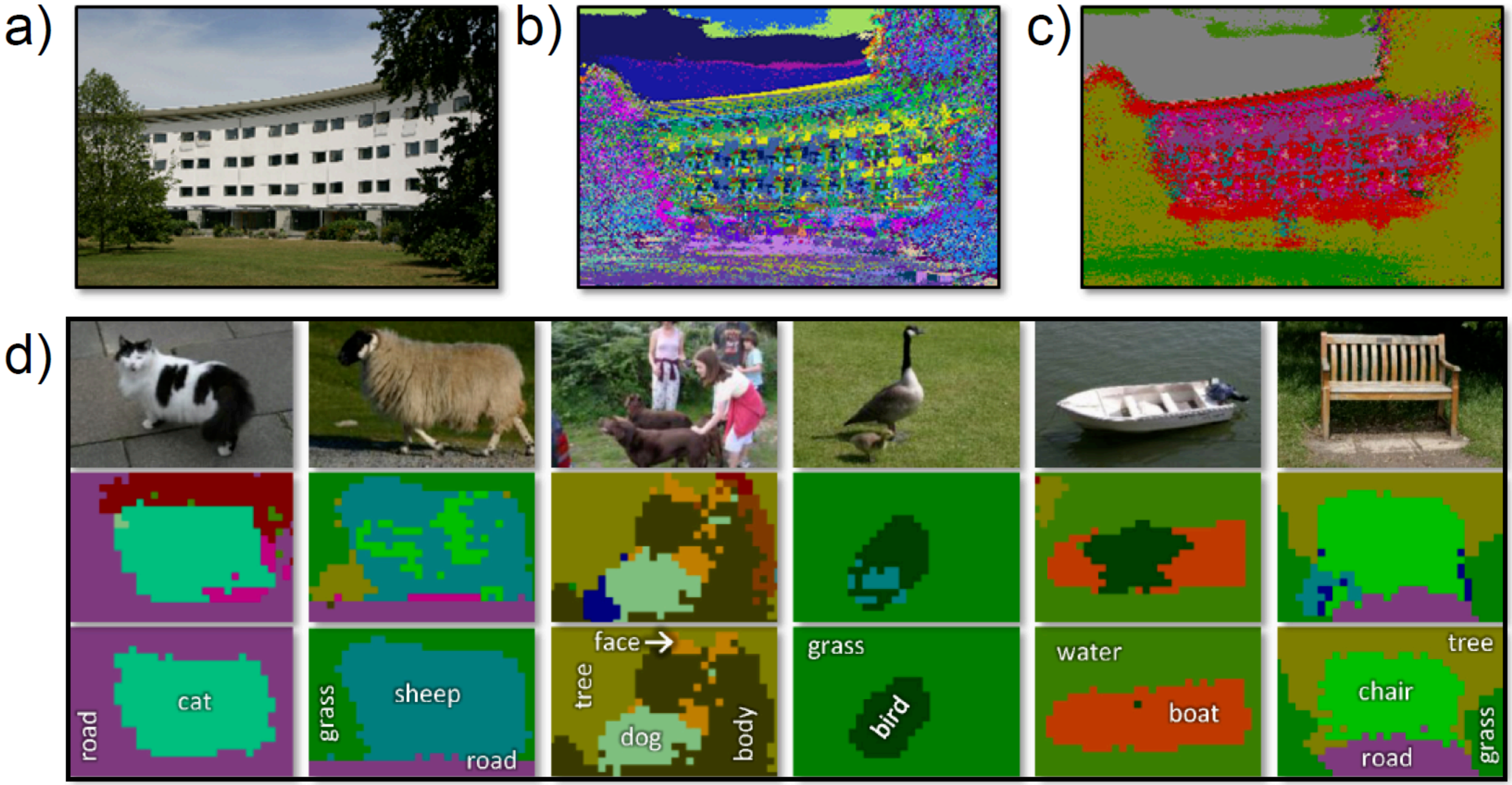
a multi-class decision, but can be turned into a set of binary “1-vs-All” classification decisions:

person A vs everyone else
person B vs everyone else
person C vs everyone else
...

Classification Example

Semantic Segmentation

assign semantic label to each pixel by combining several 1-vs-All binary classifiers



Ambiguity of visual world

- Unfortunately visual measurements may be compatible with more than one world state **w**
  - Measurement process is noisy
  - Inherent ambiguity in visual data
- Conclusion: the best we can do is compute a probability distribution Pr(**w**|**x**) over possible states of world

Refined goal of computer vision

- Take observations **x**
- Return probability distribution Pr(**w**|**x**) over possible worlds compatible with data

(not always tractable – might have to settle for an approximation to this distribution, samples from it, or the best (MAP) solution for **w**)

Components of solution

We need

- A model that mathematically relates the visual data $\mathbf{x}$ to the world state $\mathbf{w}$. Model specifies family of relationships, particular relationship depends on parameters θ

- A learning algorithm: fits parameters θ from paired training examples $\mathbf{x}_i,\mathbf{w}_i$

- An inference algorithm: uses model to return $Pr(\mathbf{w}|\mathbf{x})$ given new observed data $\mathbf{x}$.

Types of Model

The model mathematically relates the visual data $\mathbf{x}$ to the world state $\mathbf{w}$. Two main categories of model

1. Model contingency of the world on the data $Pr(\mathbf{w}|\mathbf{x})$
2. Model contingency of data on world $Pr(\mathbf{x}|\mathbf{w})$

Generative vs. Discriminative

1. Model contingency of the world on data Pr(**w**|**x**)
   (DISCRIMINATIVE MODEL)

2. Model contingency of data on world Pr(**x**|**w**)
   (GENERATIVE MODELS)

   called “Generative” because when we draw samples from the model, we GENERATE new data

Type 1: Model Pr(**w**|**x**) - Discriminative

How to model Pr(**w**|**x**)?

1. Choose an appropriate form for Pr(**w**)
2. Make parameters a function of **x**
3. Function takes parameters θ that define its shape

Learning algorithm: learn parameters θ from training data **x**,**w**

Inference algorithm: just evaluate Pr(**w**|**x**)

Type 2: Pr(**x**|**w**) - Generative

How to model Pr(**x**|**w**)?

1. Choose an appropriate form for Pr(**x**)
2. Make parameters a function of **w**
3. Function takes parameters θ that define its shape

Learning algorithm: learn parameters θ from training data **x**,**w**

Inference algorithm: Define prior Pr(**w**) and then compute Pr(**w**|**x**) using Bayes' rule

$$Pr(\mathbf{w}|\mathbf{x}) = \frac{Pr(\mathbf{x}|\mathbf{w})Pr(\mathbf{w})}{\int Pr(\mathbf{x}|\mathbf{w})Pr(\mathbf{w})d\mathbf{w}}$$

Summary

Two different types of model depend on the quantity of interest:

1. **Pr(w|x)** Discriminative
2. **Pr(w|x)** Generative

Inference in discriminative models easy as we directly model posterior Pr(**w**|**x**). Generative models require more complex inference process using Bayes' rule

Outline

- Computer vision models
  - Discriminative vs generative
- Worked example 1: Regression
- Worked example 2: Classification
- Which type should we choose?
- Applications

Worked example 1: Regression

Consider simple case where

- we make a univariate continuous measurement **x**
- use this to predict a univariate continuous state **w**

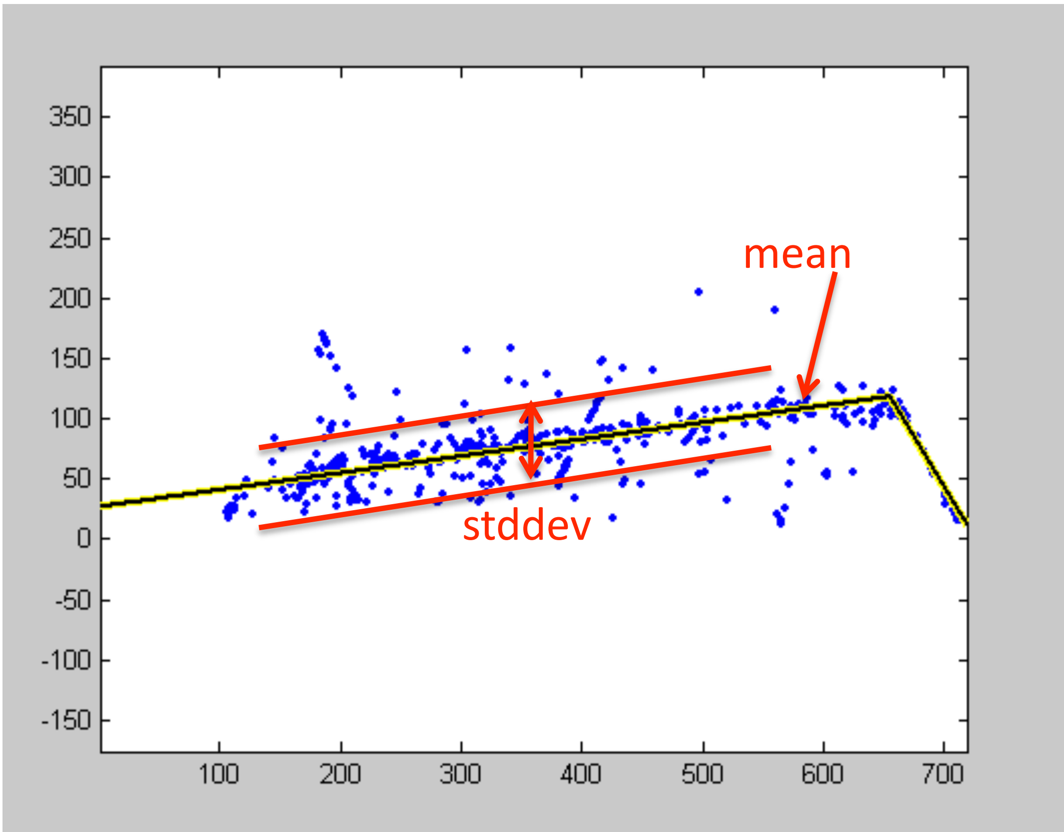
(recall, when world state is continuous we call
our inferencing procedure "regression")

Sample Regression Application

Example: learning bounding box height distribution as a function of image row

W (“world”)

half height of person in pixels



row number

X (data)

Ge and Collins, CVPR 2009
“Marked PointProcesses
for Crowd Counting”

Type 1: Model Pr(**w**|**x**) - Discriminative

How to model Pr(**w**|**x**)?

1. Choose an appropriate form for Pr(**w**)
2. Make parameters a function of **x**
3. Function takes parameters θ that define its shape

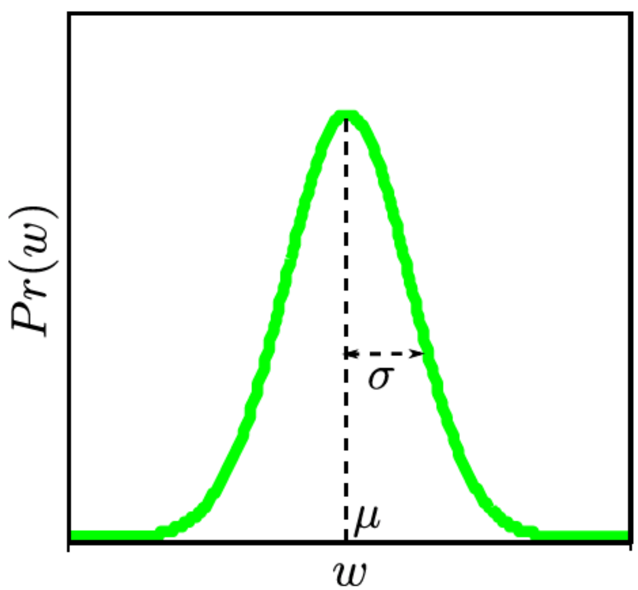
Learning algorithm: learn parameters θ from training data **x**,**w**

Inference algorithm: just evaluate Pr(**w**|**x**)

Type 1: Model Pr(**w**|**x**) - Discriminative

How to model Pr(**w**|**x**)?

1. Choose an appropriate form for Pr(**w**)
2. Make parameters a function of **x**
3. Function takes parameters θ that define its shape



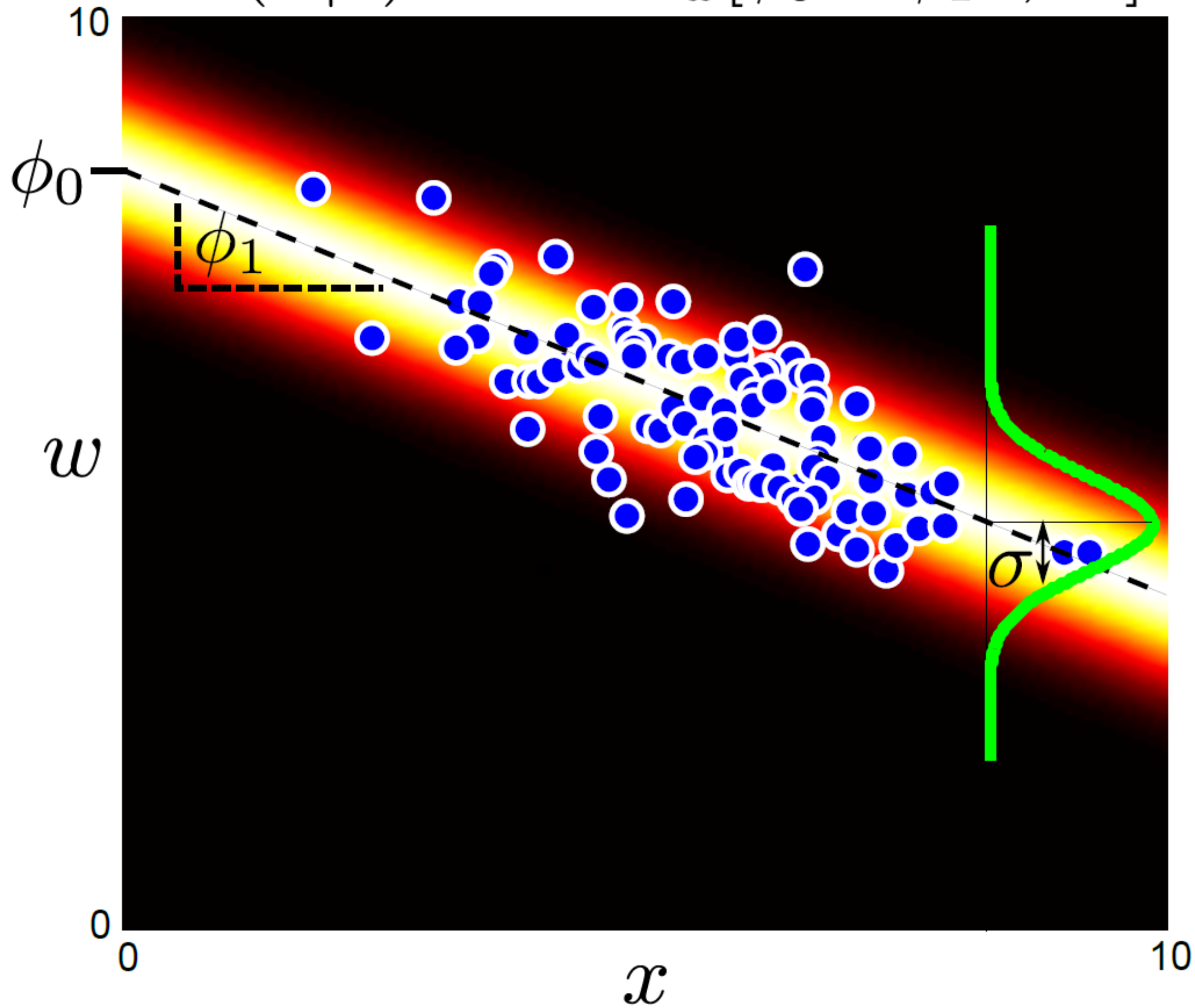
1. Choose normal distribution over w
2. Make mean μ a linear function of x
   Let variance be a constant

$$Pr(w|x, \boldsymbol{\theta}) = \text{Norm}_w\left[\phi_0 + \phi_1 x, \sigma^2\right]$$

3. Model parameters are ϕ_0, ϕ_1, σ^2.

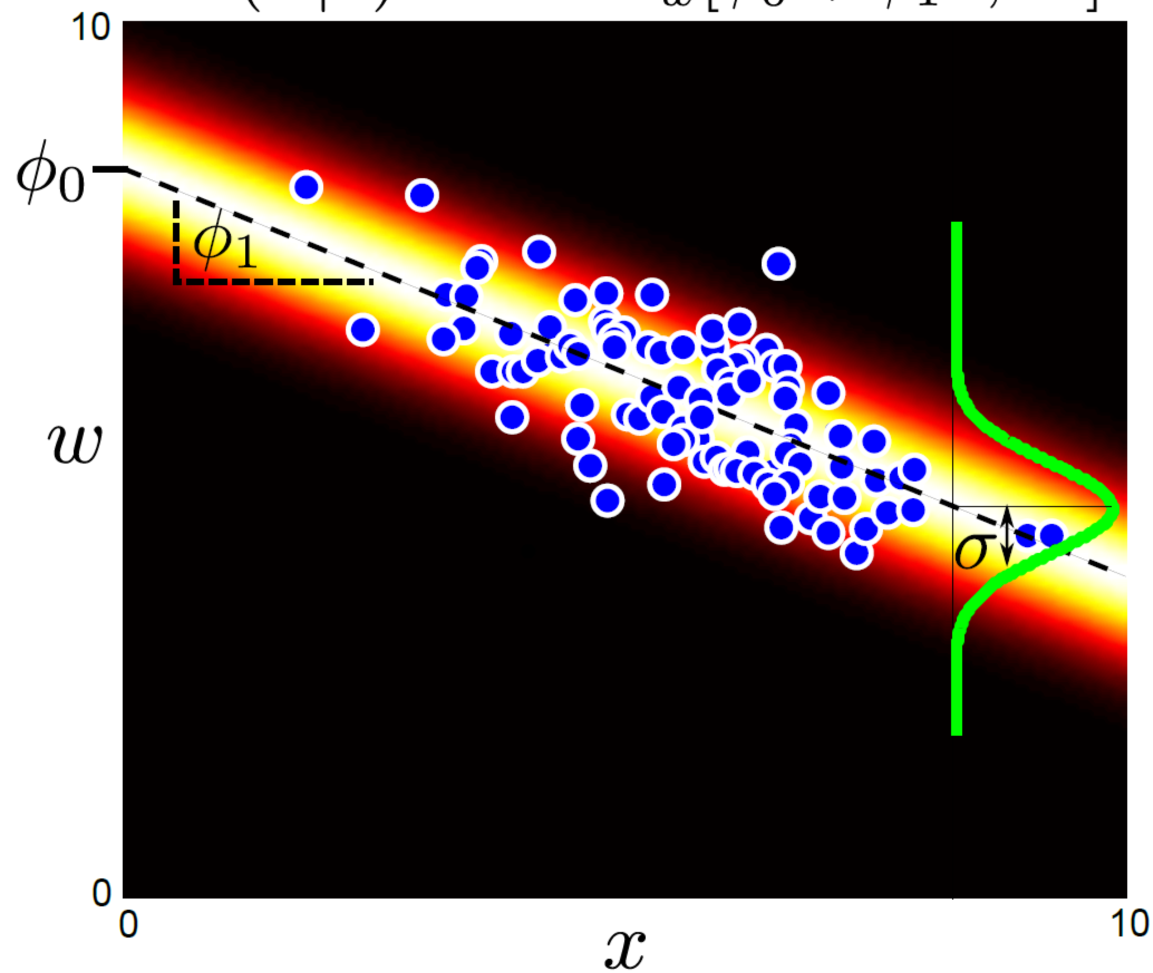
note: This is a *linear regression* model.

$$Pr(w|x) = \text{Norm}_w[\phi_0 + \phi_1 x, \sigma^2]$$



Parameters $\boldsymbol{\theta} = \{\phi_0, \phi_1, \sigma^2\}$ are y-offset, slope and variance

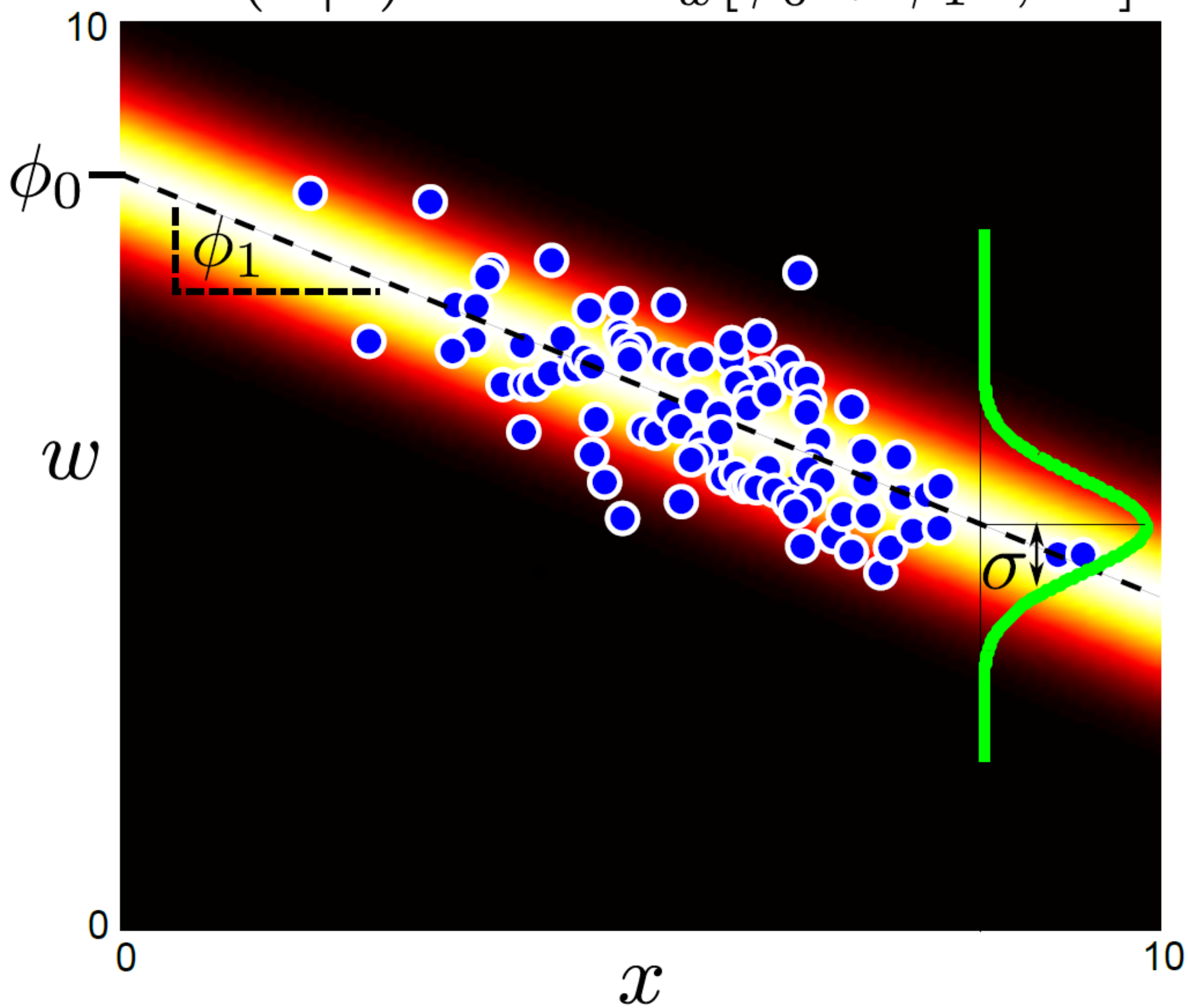
$$Pr(w|x) = \text{Norm}_w[\phi_0 + \phi_1 x, \sigma^2]$$



Learning algorithm: learn θ from training pairs (x_i,w_i). E.g. MAP

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \underset{\boldsymbol{\theta}}{\arg\max}\, Pr(\boldsymbol{\theta}|w_{1\ldots I}, x_{1\ldots I}) \\ &= \underset{\boldsymbol{\theta}}{\arg\max}\, Pr(w_{1\ldots I}|x_{1\ldots I}, \boldsymbol{\theta})Pr(\boldsymbol{\theta}) \quad = \underset{\boldsymbol{\theta}}{\arg\max} \prod_{i=1}^{I} Pr(w_i|x_i, \boldsymbol{\theta})Pr(\boldsymbol{\theta}),\end{aligned}$$

$$Pr(w|x) = \text{Norm}_w[\phi_0 + \phi_1 x, \sigma^2]$$



Inference algorithm: just evaluate Pr(**w**|**x**) for new data **x**

note: this gives us a whole normal distribution over w, but we could then use the mean to make a prediction.

Type 2: Pr(**x**|**w**) - Generative

How to model Pr(**x**|**w**)?

1. Choose an appropriate form for Pr(**x**)
2. Make parameters a function of **w**
3. Function takes parameters θ that define its shape

Learning algorithm: learn parameters θ from training data **x**,**w**

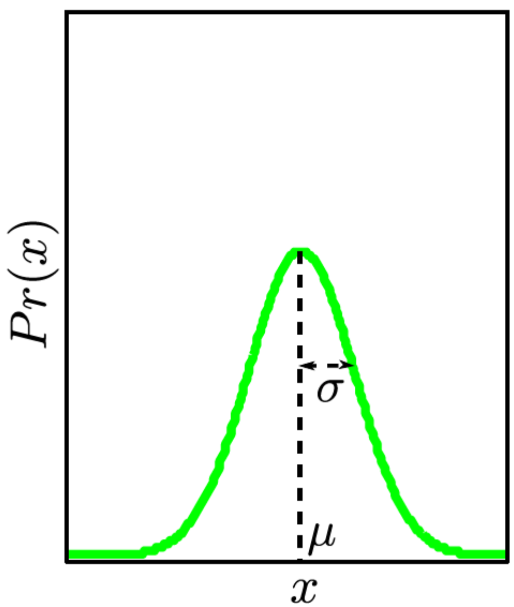
Inference algorithm: Define prior Pr(**w**) and compute Pr(**w**|**x**) using Bayes' rule

$$Pr(\mathbf{w}|\mathbf{x}) = \frac{Pr(\mathbf{x}|\mathbf{w})Pr(\mathbf{w})}{\int Pr(\mathbf{x}|\mathbf{w})Pr(\mathbf{w})d\mathbf{w}}$$

Type 2: Pr(**x**|**w**) - Generative

How to model Pr(**x**|**w**)?

1. Choose an appropriate form for Pr(**x**)
2. Make parameters a function of **w**
3. Function takes parameters θ that define its shape

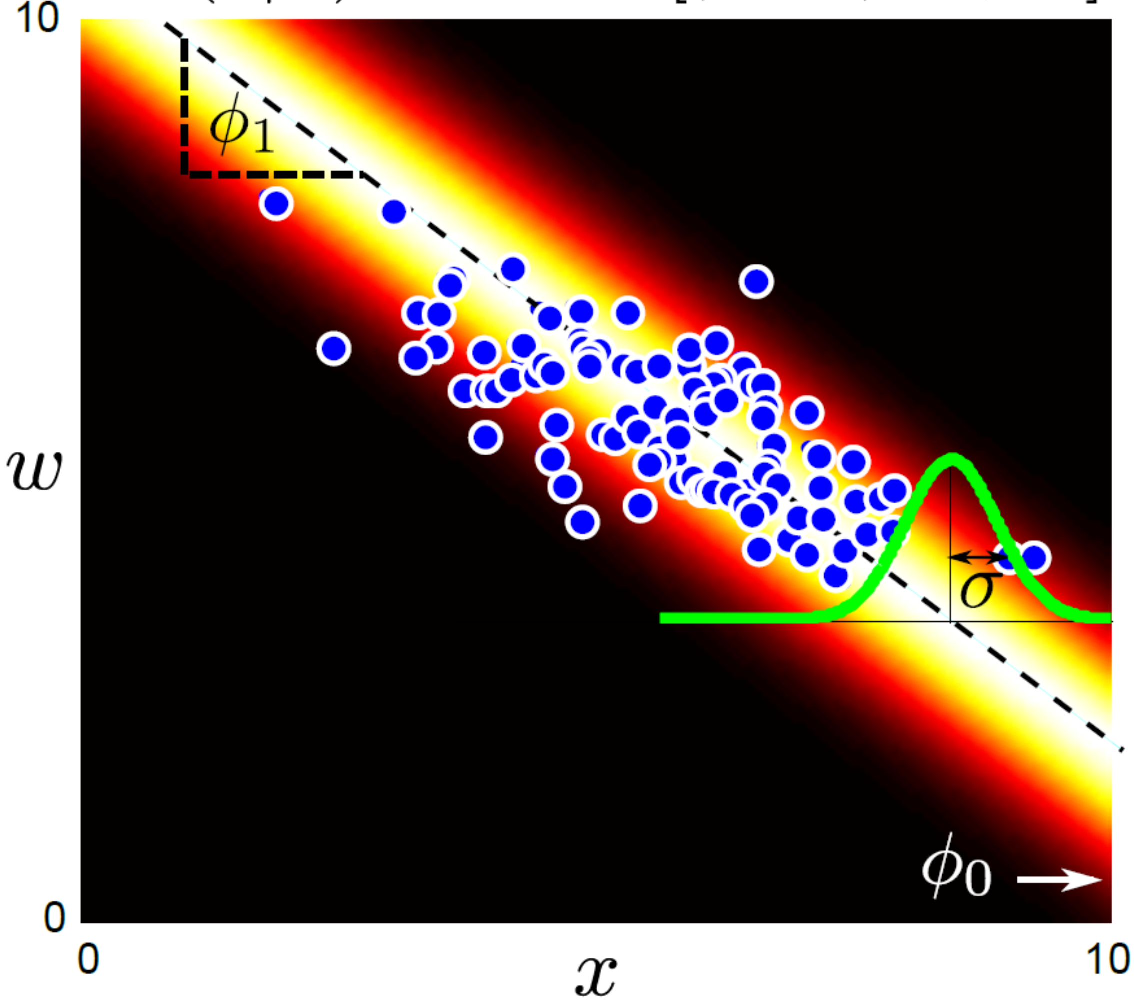


1. Choose normal distribution over x
2. Make mean μ linear function of w (variance constant)

$$Pr(x|w,\boldsymbol{\theta}) = \text{Norm}_x\left[\phi_0 + \phi_1 w, \sigma^2\right]$$

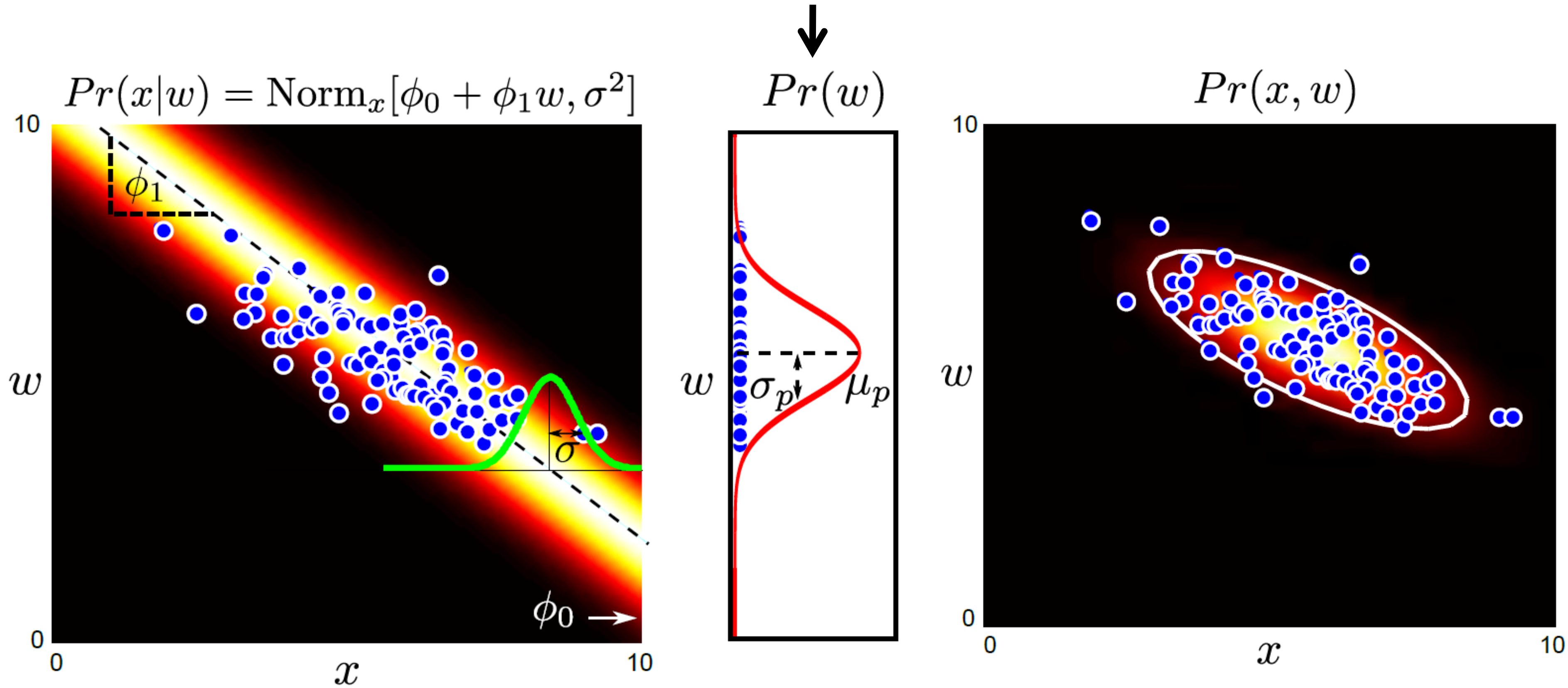
3. Parameter are ϕ_0, ϕ_1, σ^2.

$$Pr(x|w) = \text{Norm}_x[\phi_0 + \phi_1 w, \sigma^2]$$



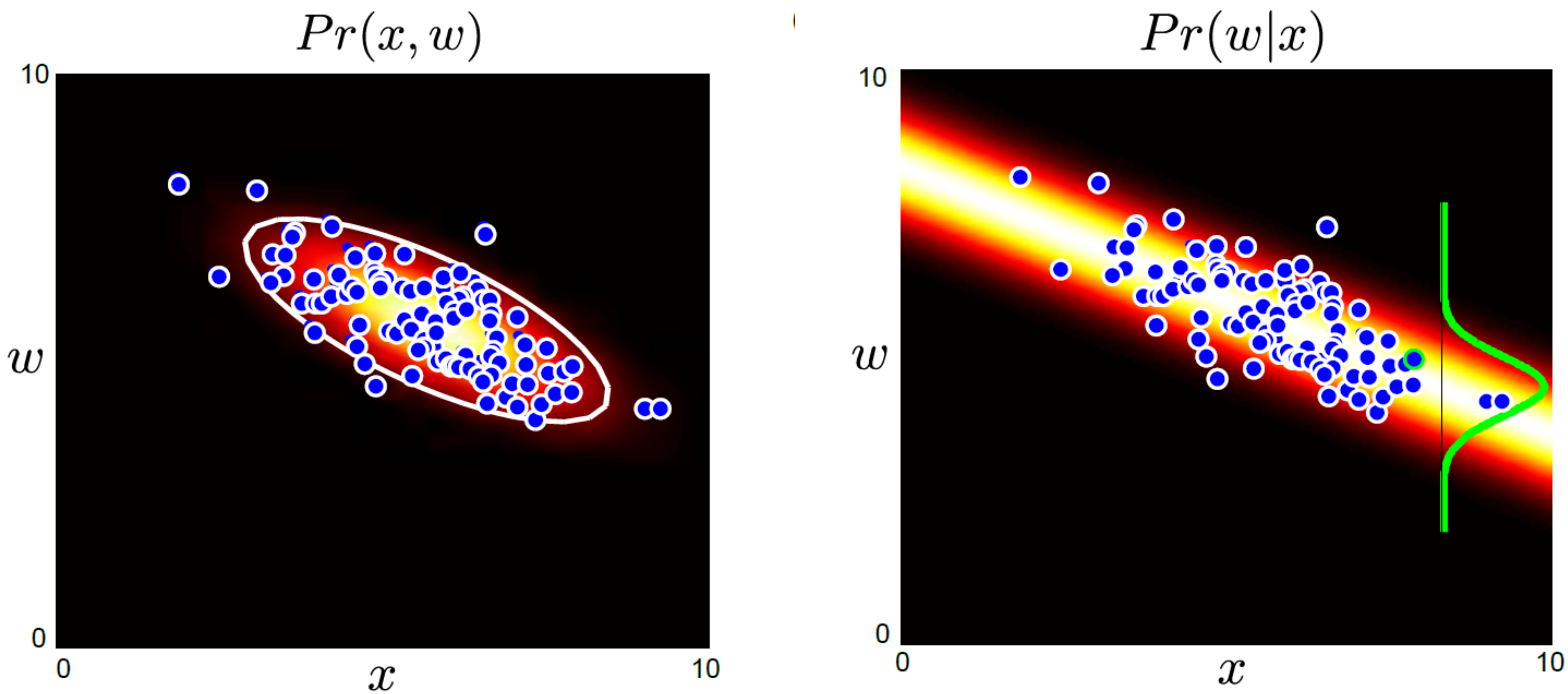
Learning algorithm: learn θ from training pairs (x_i,w_i).

can also learn prior distribution from the w_i components of the training data pairs, or specify in some other way



$Pr(x|w)$ x $Pr(w)$ = $Pr(x,w)$

joint probability



Inference algorithm: compute Pr(**w**|**x**) using Bayes rule

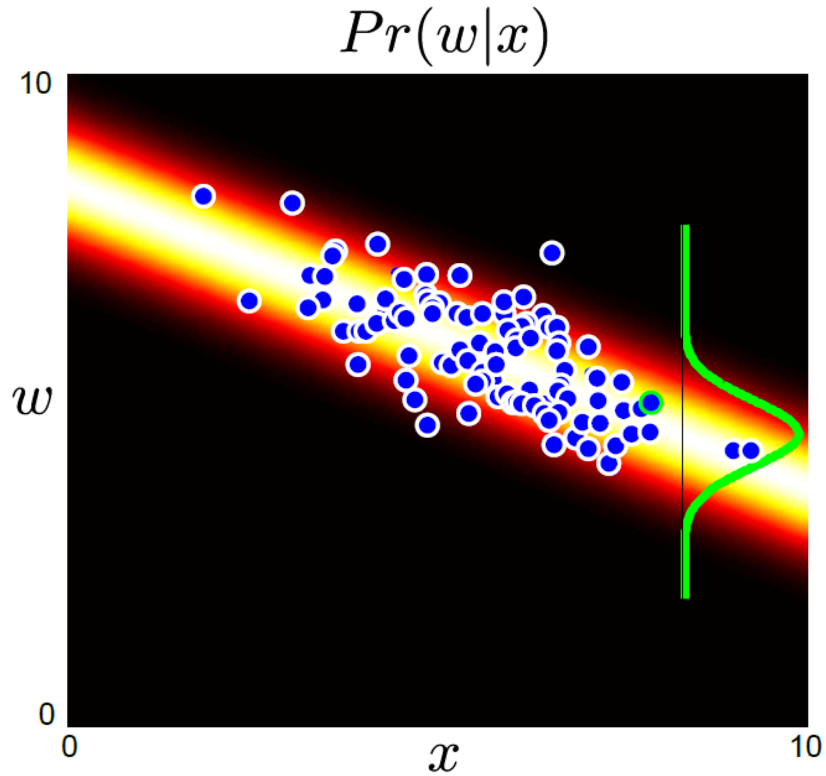
$$Pr(\mathbf{w}|\mathbf{x}) = \frac{Pr(\mathbf{x}|\mathbf{w})Pr(\mathbf{w})}{\int Pr(\mathbf{x}|\mathbf{w})Pr(\mathbf{w})d\mathbf{w}}$$

Interesting Observation

In this example, both generative and discriminative models lead to the same posterior normal distribution p(w|x) if MLE is used to estimate model parameters.

This is due to:
* both x and w being continuous
* they are related by a linear model
* normal distributions were to represent all uncertainties.

MAP estimation would have led to differences in the generative and discriminative results because we would place priors on the parameters (e.g. ϕ_0, ϕ_1, σ^2) and the parameters have different meanings in the two models.



Outline

- Computer vision models
  - Discriminative vs generative
- Worked example 1: Regression
- Worked example 2: Classification
- Which type should we choose?
- Applications

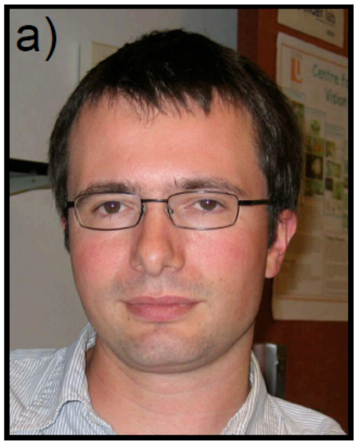
Worked example 2: Classification

Consider simple case where

- we make a univariate continuous measurement x
- use this to predict a discrete binary world w $\in \{0, 1\}$

(recall: we are calling inference “classification” when the world state is discrete)

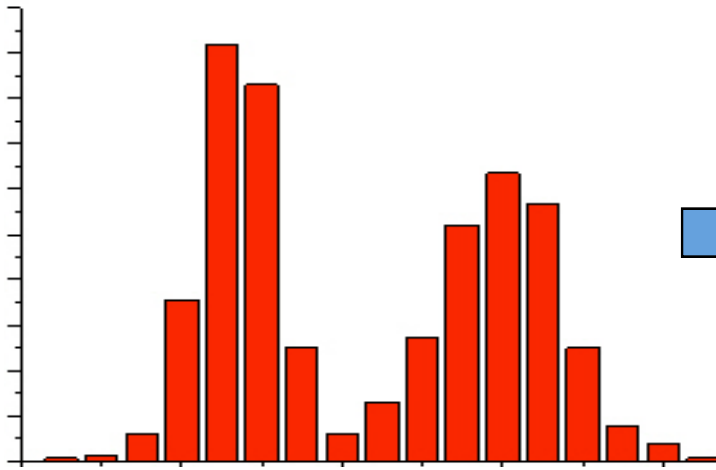
Classification Example: Skin Detection



input RGB image

+

red channel histogram



nonskin skin

offline learning from labeled training data

classify each pixel as skin vs not skin

Type 1: Model Pr(**w**|**x**) - Discriminative

How to model Pr(**w**|**x**)?

- Choose an appropriate form for Pr(**w**)
- Make parameters a function of **x**
- Function takes parameters θ that define its shape

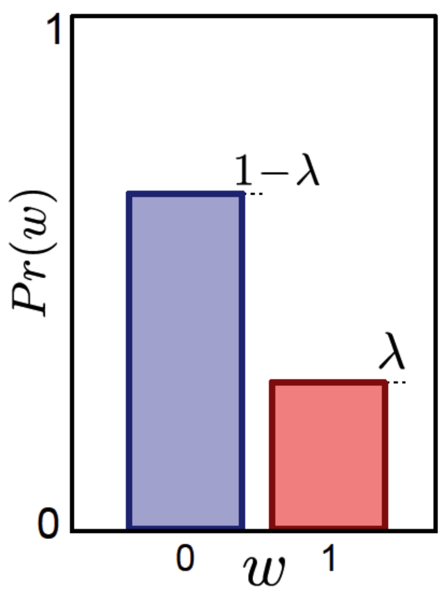
Learning algorithm: learn parameters θ from training data **x**,**w**

Inference algorithm: just evaluate **Pr(w|x)**

Type 1: Model Pr(**w**|**x**) - Discriminative

How to model Pr(**w**|**x**)?

1. Choose an appropriate form for Pr(**w**)
2. Make parameters a function of **x**
3. Function takes parameters $\boldsymbol{\theta}$ that define its shape



1. Choose Bernoulli dist. for Pr(**w**)
2. Make parameter λ a function of **x**

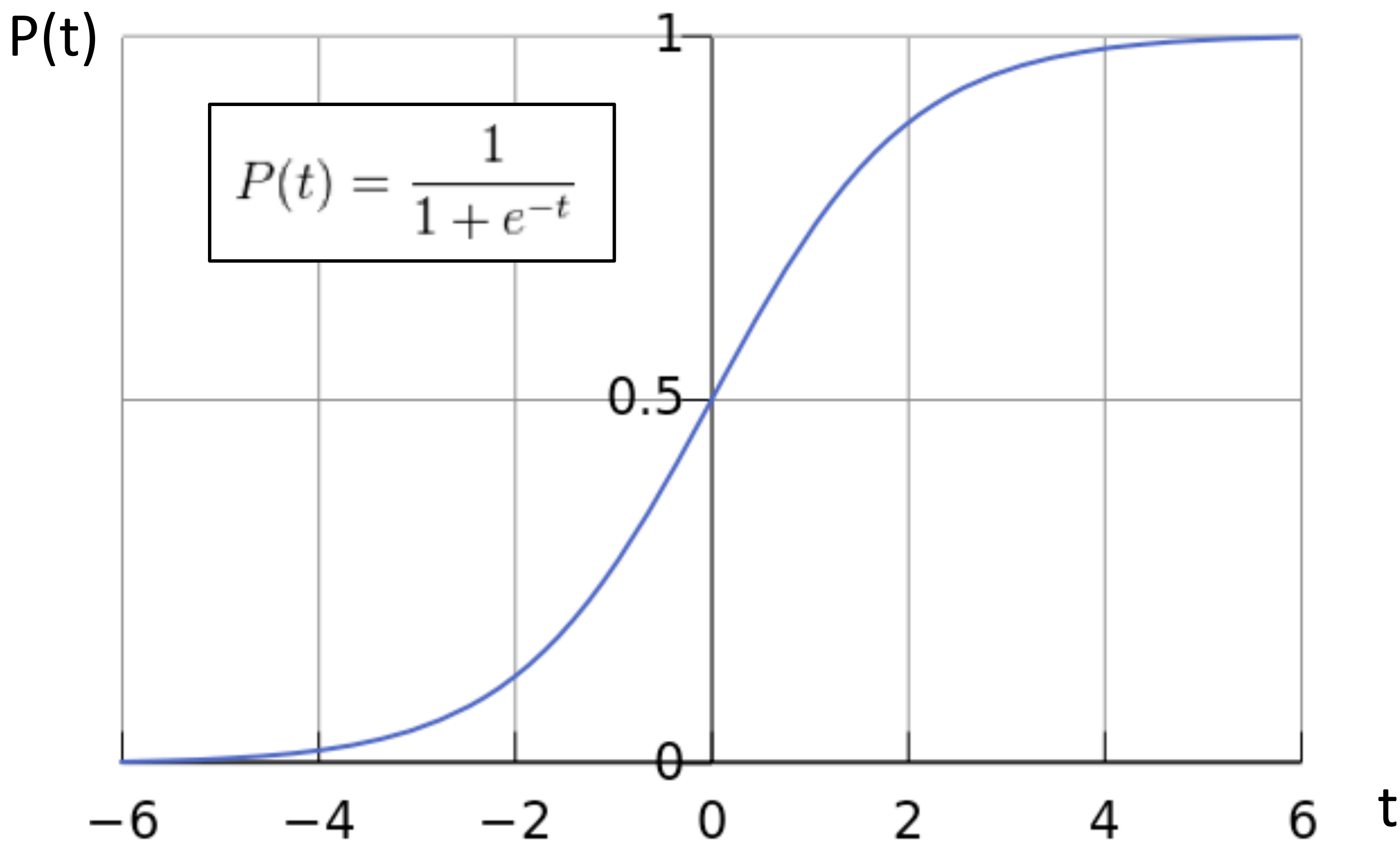
$$
\begin{aligned}
Pr(w|x) &= \text{Bern}_w\left[\text{sig}[\phi_0 + \phi_1 x]\right] \\
&= \text{Bern}_w\left[\frac{1}{1+\exp[-\phi_0 - \phi_1 x]}\right]
\end{aligned}
$$

3. Function takes parameters ϕ_0 and ϕ_1

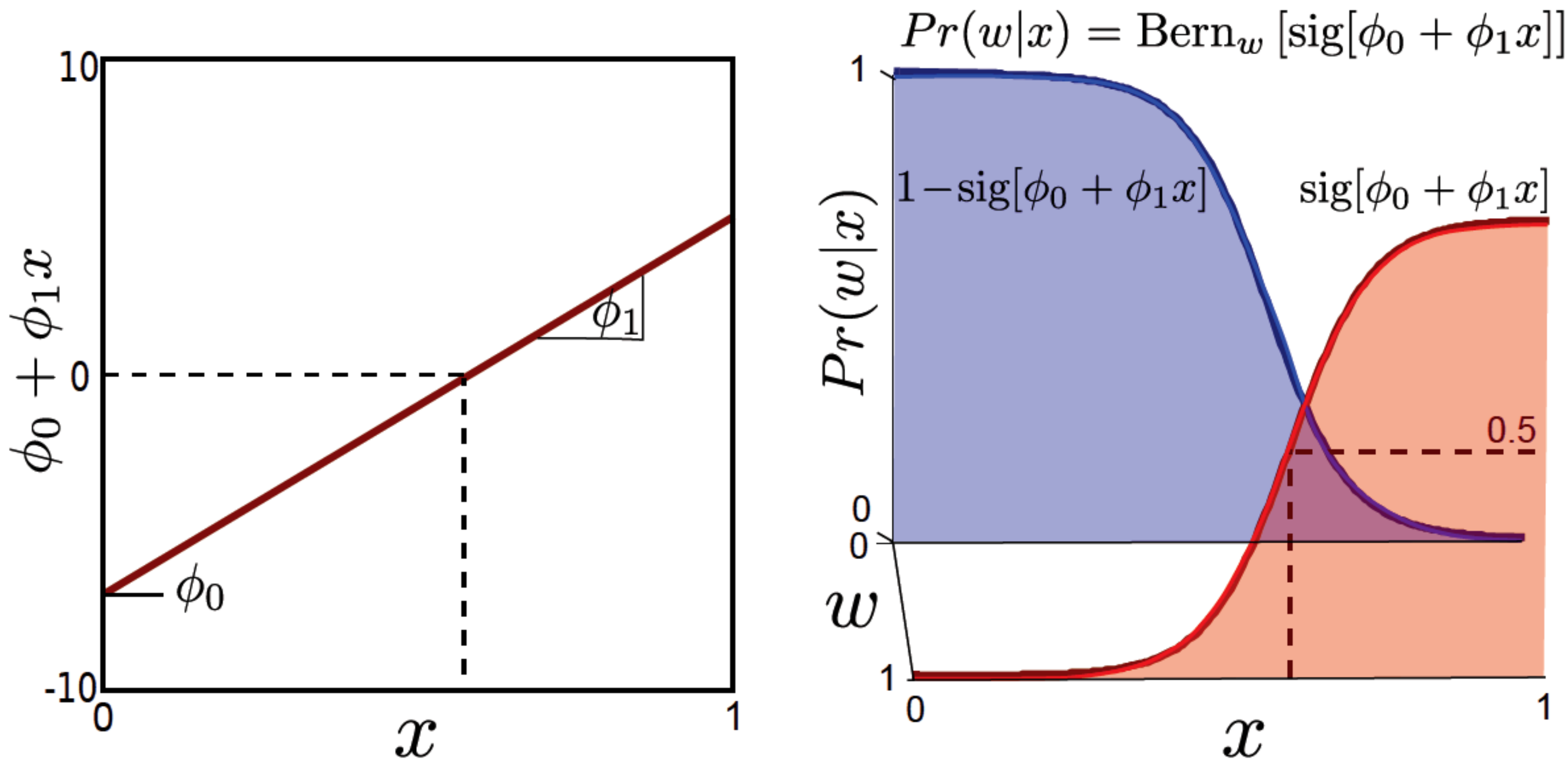
note: This model is called *logistic regression (even though we are doing classification here not regression)*

Computer vision: models, learning and inference. ©2011 Simon J.D. Prince

Background: Logistic or “sig” Function



Maps real number line into range [0,1]



Learn the two model parameters $\boldsymbol{\theta} = \{\phi_0, \phi_1\}$
from training pairs (x_i,w_i) by standard methods (ML,MAP, Bayesian)

Inference: Just evaluate Pr(w|x)

Computer vision: models, learning and inference. ©2011 Simon J.D. Prince

Type 2: Pr(**x**|**w**) - Generative

How to model Pr(**x**|**w**)?

1. Choose an appropriate form for Pr(**x**)
2. Make parameters a function of **w**
3. Function takes parameters θ that define its shape

Learning algorithm: learn parameters θ from training data **x**,**w**

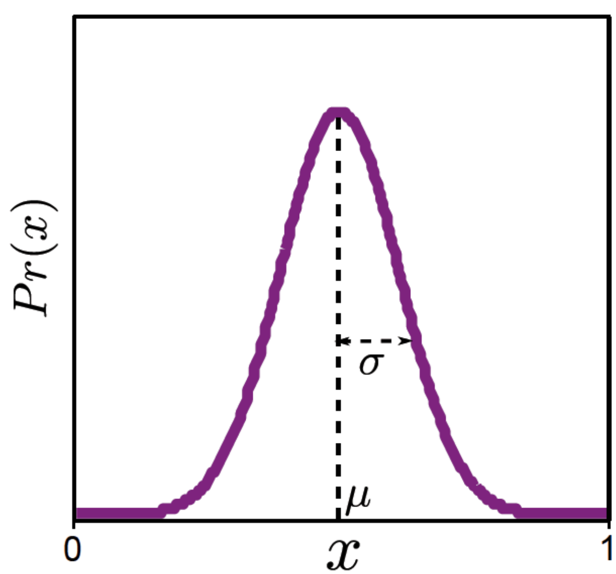
Inference algorithm: Define prior Pr(**w**) and then compute Pr(**w**|**x**) using Bayes' rule

$$Pr(\mathbf{w}|\mathbf{x}) = \frac{Pr(\mathbf{x}|\mathbf{w})Pr(\mathbf{w})}{\int Pr(\mathbf{x}|\mathbf{w})Pr(\mathbf{w})d\mathbf{w}}$$

Type 2: Pr(**x**|**w**) - Generative

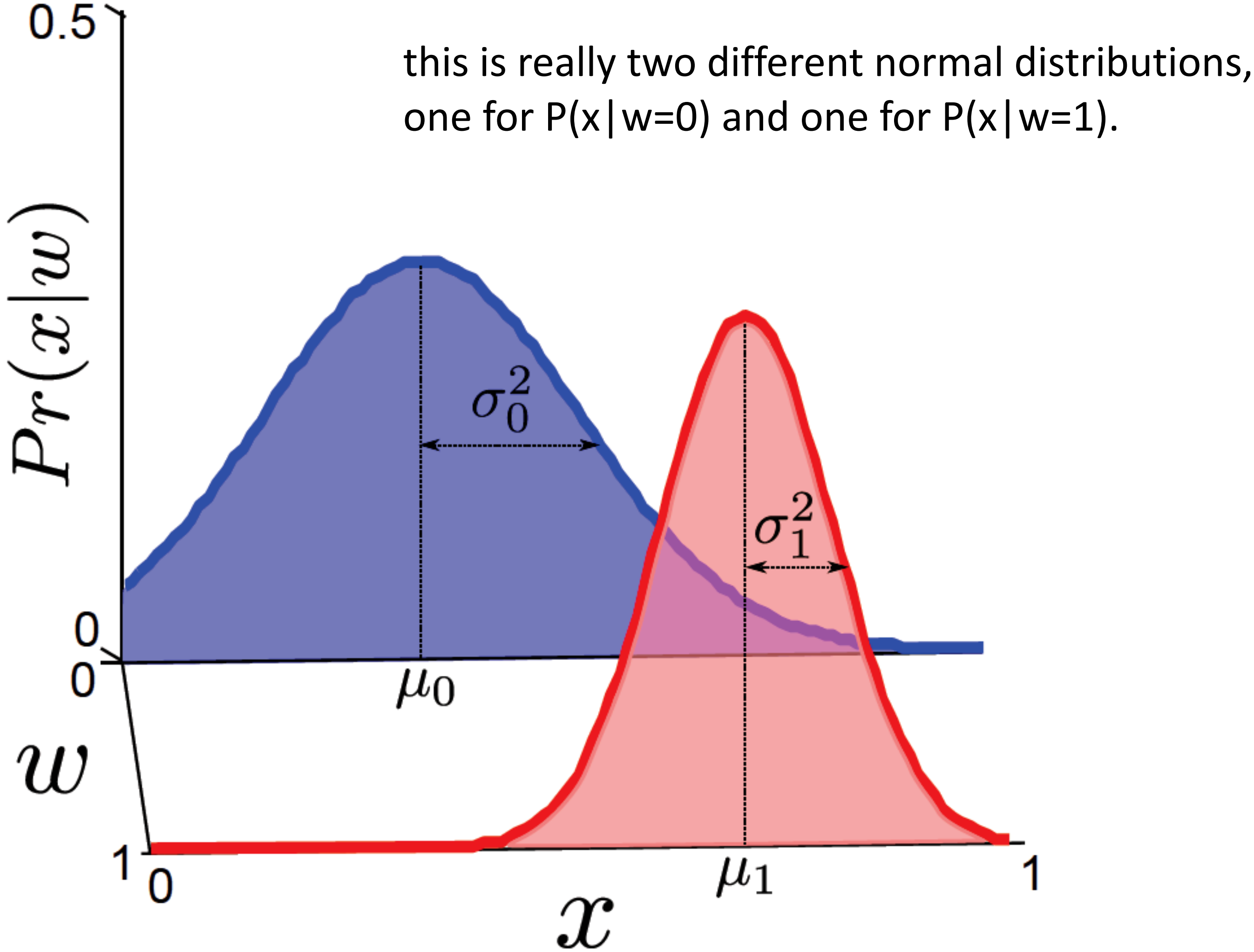
How to model Pr(**x**|**w**)?

1. Choose an appropriate form for Pr(**x**)
2. Make parameters a function of **w**
3. Function takes parameters $\boldsymbol{\theta}$ that define its shape



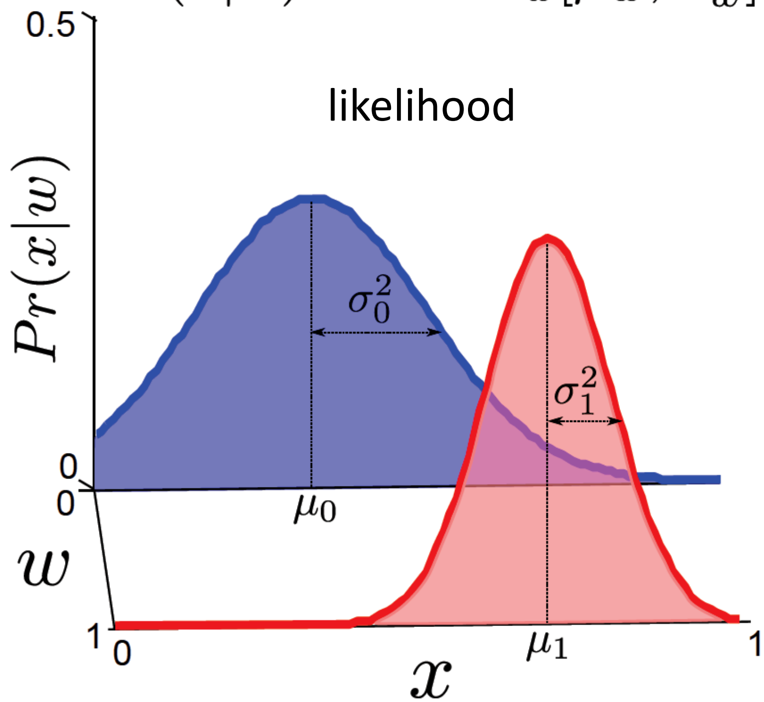
1. Choose a Gaussian distribution for Pr(**x**)
2. Make parameters a function of discrete binary **w**
$$Pr(x|w) = \text{Norm}_x[\mu_w, \sigma_w^2]$$
3. Function takes parameters μ_0, μ_1, $\sigma^2{}_0$, $\sigma^2{}_1$ that define its shape

$$Pr(x|w) = \text{Norm}_x[\mu_w, \sigma^2_w]$$

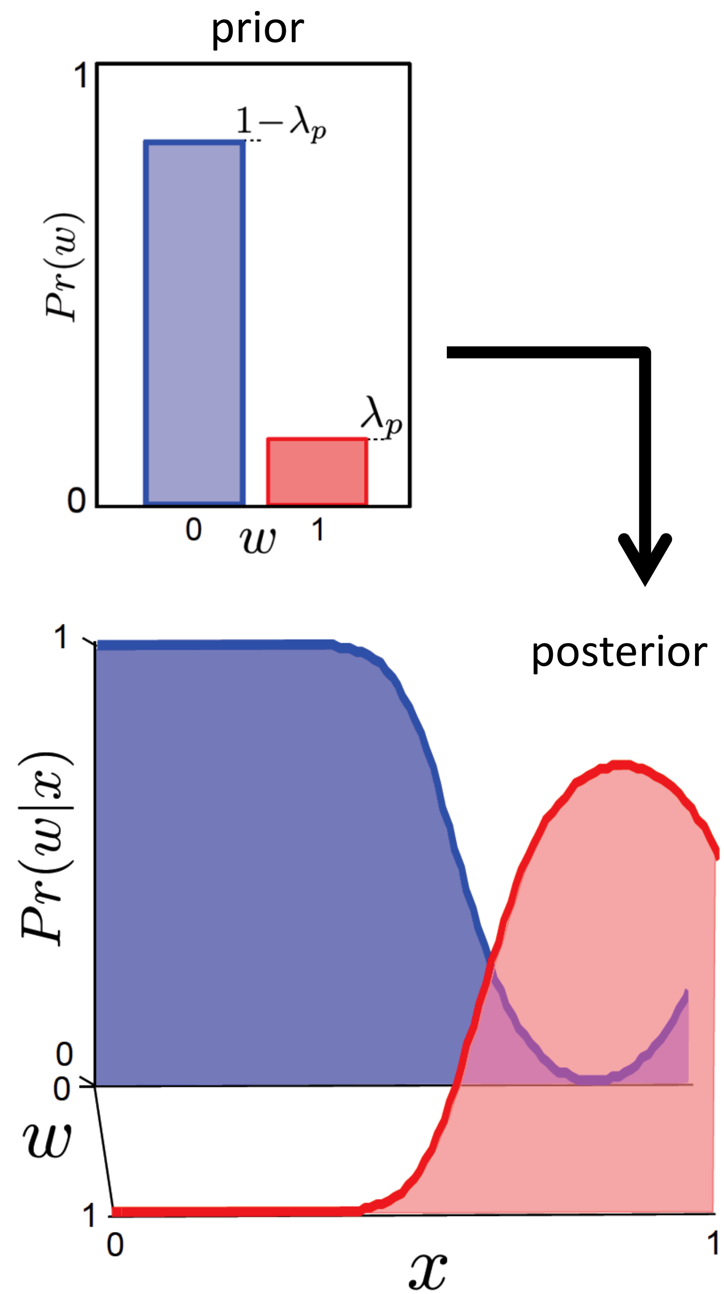


Learn model parameters μ_0, μ_1, σ^2_0, σ^2_1 from training data

$$Pr(x|w) = \text{Norm}_x[\mu_w, \sigma_w^2]$$



*

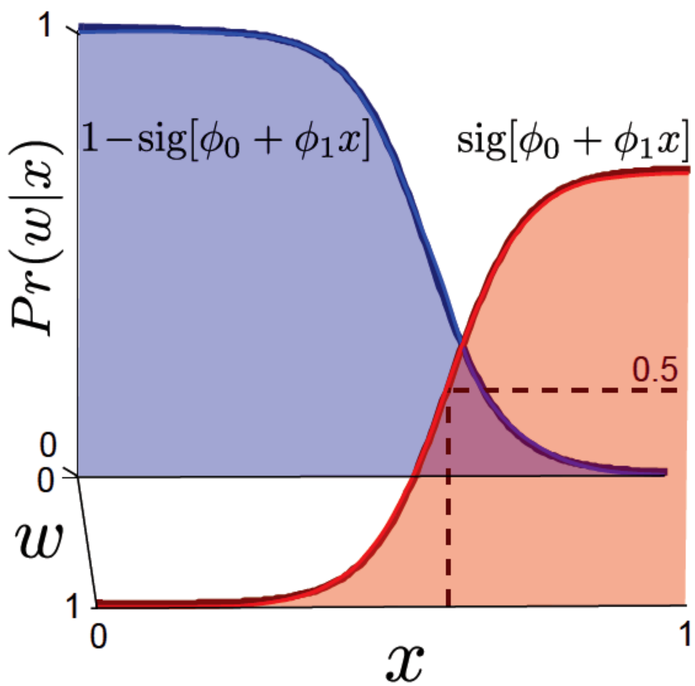


Inference algorithm: Define prior Pr(**w**) and then compute Pr(**w**|**x**) using Bayes' rule

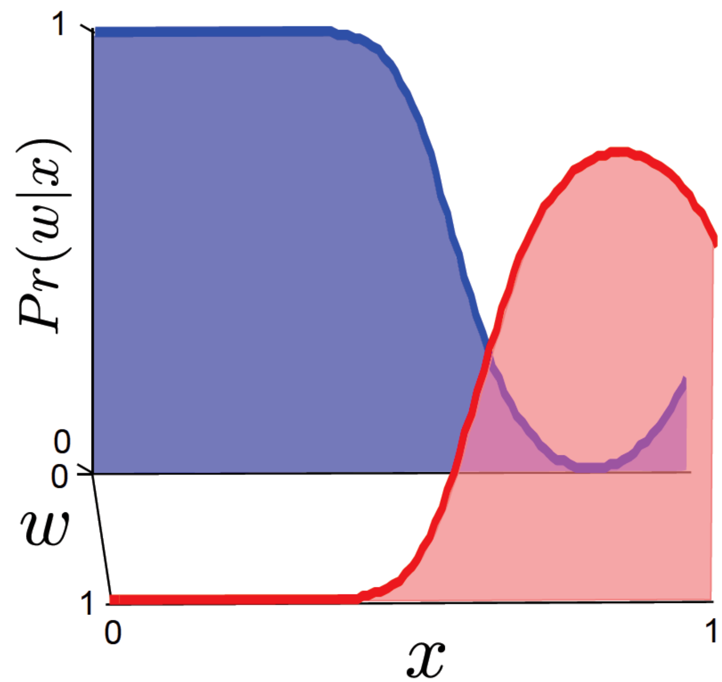
$$Pr(\mathbf{w}|\mathbf{x}) = \frac{Pr(\mathbf{x}|\mathbf{w})Pr(\mathbf{w})}{\int Pr(\mathbf{x}|\mathbf{w})Pr(\mathbf{w})d\mathbf{w}}$$

Comparing Posteriors

discriminative



generative



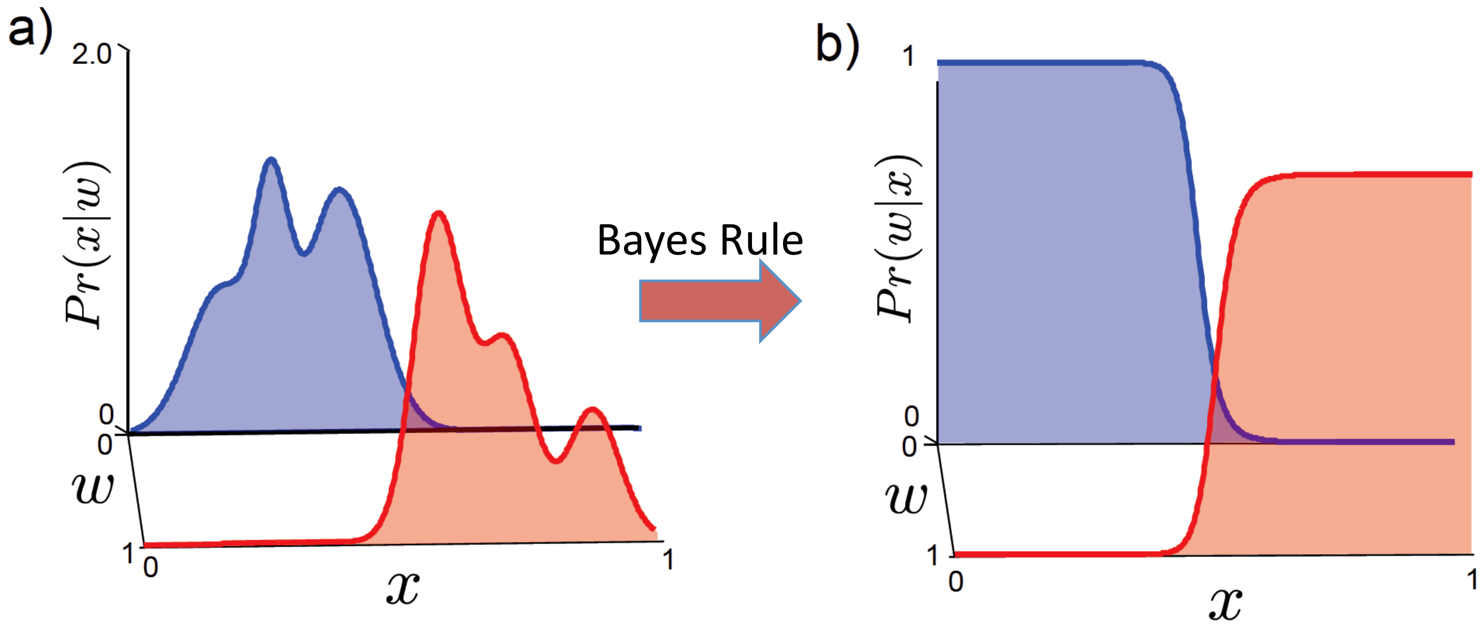
This time the posteriors are not equivalent. This is partly due to the "asymmetry" between world state (discrete data) and measurements (continuous data). Also, the shapes of discriminative posteriors tend by definition to be simpler than generative ones.

Outline

- Computer vision models
  - Discriminative vs generative
- Worked example 1: Regression
- Worked example 2: Classification
- Which type of model should we choose?
- Applications

Which type of model to use?

1. Generative methods model data – costly and many aspects of data may have no influence on world state



need lots of data to accurately model complicated multimodal likelihood functions

might be simpler to directly model the posterior

Which type of model to use?

2. Inference tends to be simpler and faster when using discriminative models
3. Data really is generated from the world – generative modeling matches this
4. If missing data, then generative preferred
5. Generative allows imposition of prior knowledge specified by user

Conclusion

- To do computer vision we build a model relating the image data **x** to the world state that we wish to estimate **w**

- Two types of model
  - Model Pr(w|x) -- discriminative
  - Model Pr(w|x) – generative

Future Plan

- Seen two types of model

	Model $Pr(w\|x)$	Model $Pr(x\|w)$
Regression $x \in [-\infty, \infty], w \in [-\infty, \infty]$	Linear regression	Linear regression
Classification $x \in [-\infty, \infty], w \in \{0, 1\}$	Logistic regression	Probability density function

  - Probability density function
  - Linear regression
  - Logistic regression
- Next three chapters concern these models