Computer vision: models, learning and inference

Chapter 6

Learning and Inference in Vision

Structure

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

Computer vision models

- 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

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

Ambiguity of visual world

- Unfortunately visual measurements may be compatible with more than one world state $\mathbf{w}$
  – Measurement process is noisy
  – Inherent ambiguity in visual data
- Conclusion: the best we can do is compute a probability distribution $Pr(\mathbf{w}|\mathbf{x})$ 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 **x** to the world state **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 **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 the data $Pr(\mathbf{w}|\mathbf{x})$

   (DISCRIMINATIVE MODEL)

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

   (GENERATIVE MODELS)

Generative as probability model over data and so when we draw samples from 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

Structure

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

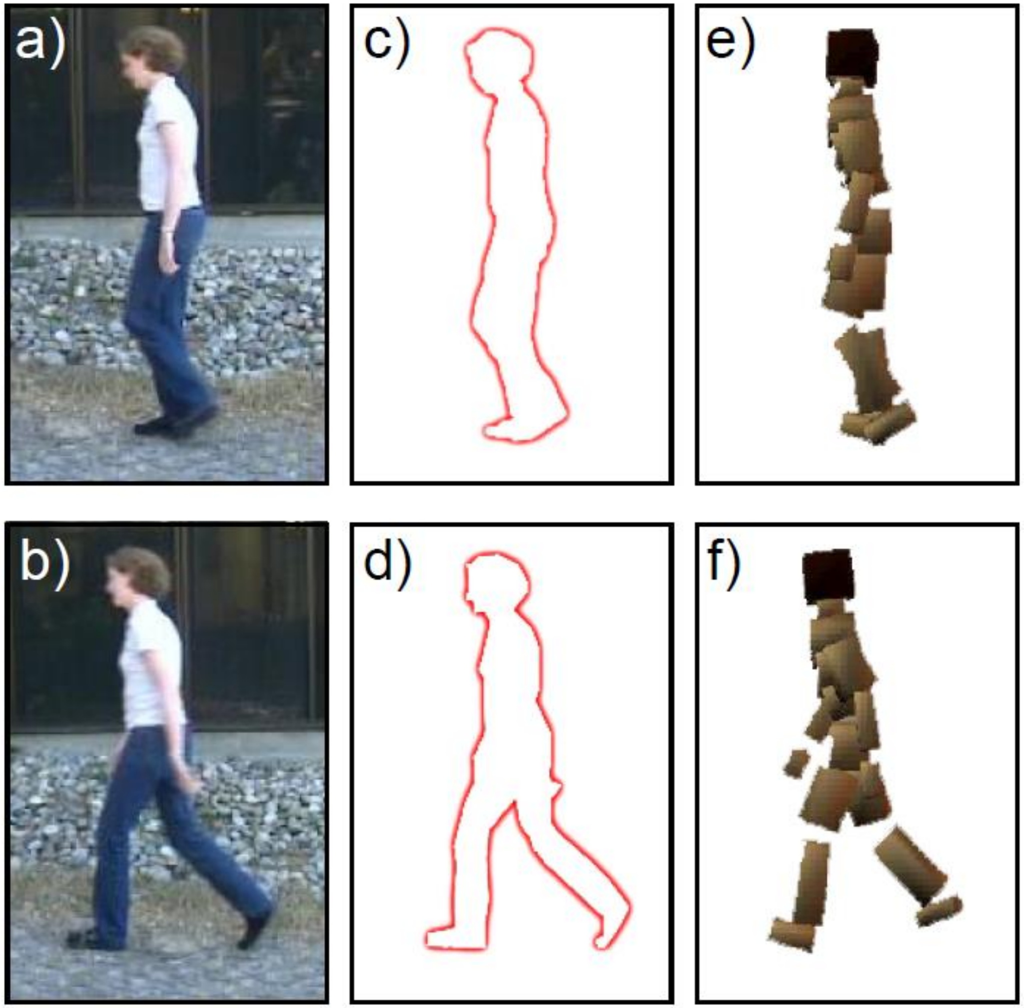
Worked example 1: Regression

Consider simple case where

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

(regression as world state is continuous)

Regression application 1: Pose from Silhouette



Regression application 2: Head pose estimation

-76^o

-11^o

2^o

8^o

43^o

79^o

Worked example 1: Regression

Consider simple case where

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

(regression as world state is continuous)

Type 1: Model Pr($\mathbf{w}|\mathbf{x}$) - Discriminative

How to model Pr($\mathbf{w}|\mathbf{x}$)?

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

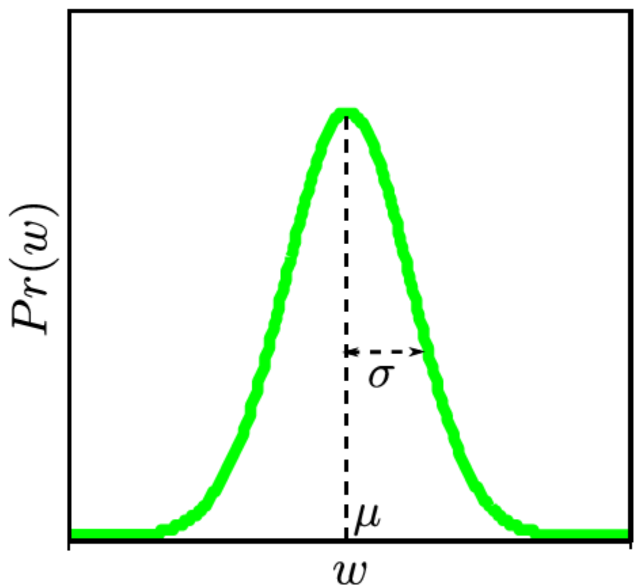
Learning algorithm: learn parameters θ from training data $\mathbf{x},\mathbf{w}$

Inference algorithm: just evaluate Pr($\mathbf{w}|\mathbf{x}$)

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



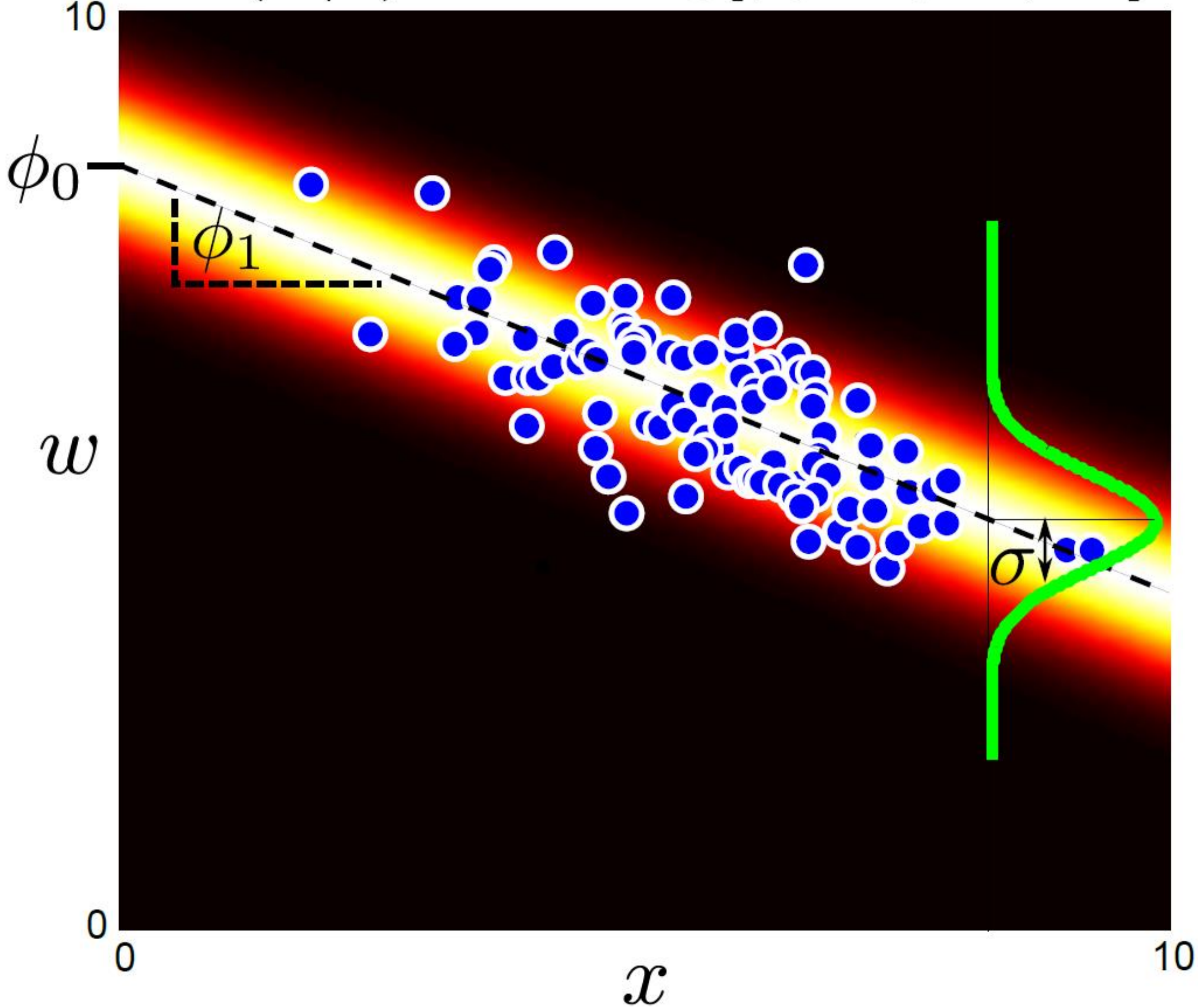
1. Choose normal distribution over w
2. Make mean μ linear function of x (variance constant)

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

3. Parameters are ϕ_0, ϕ_1, σ^2.

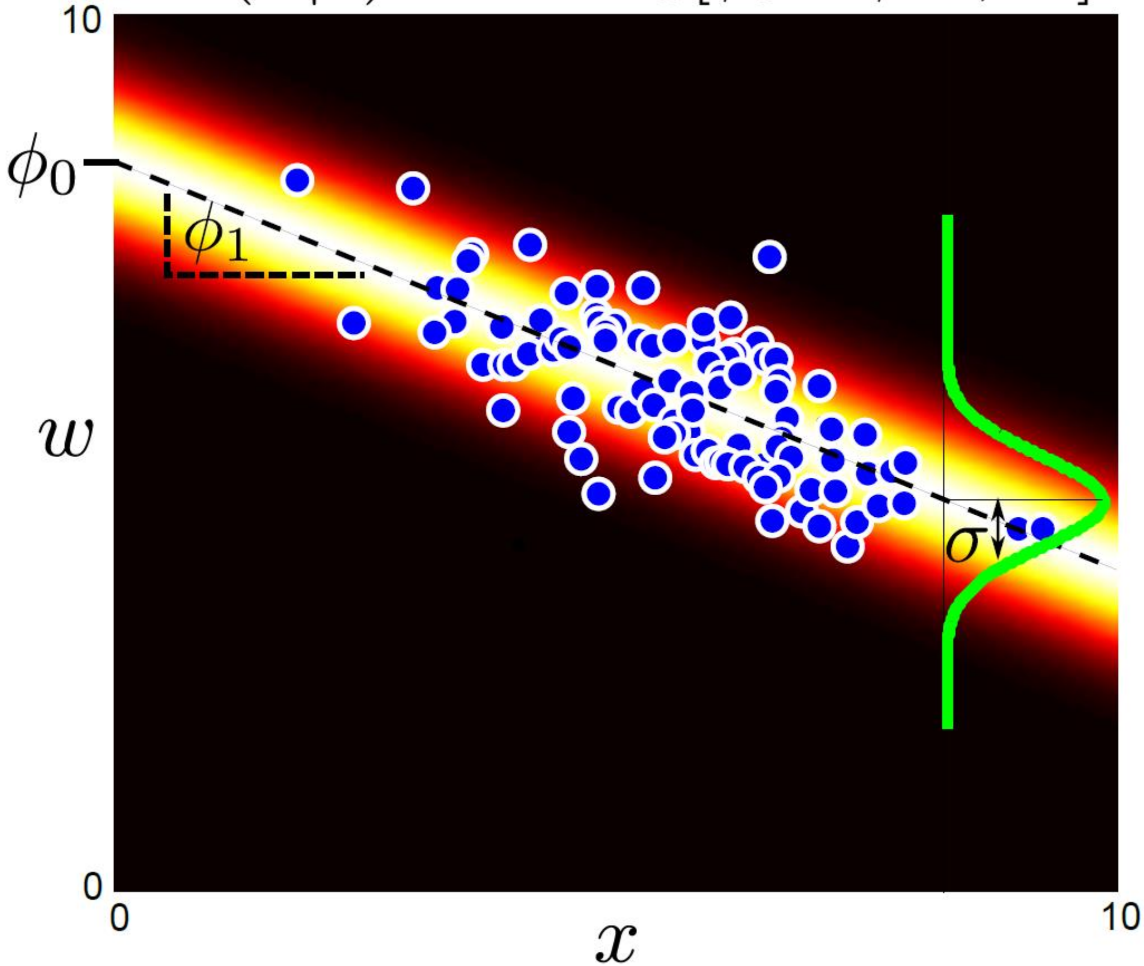
This model is called *linear regression.*

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



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

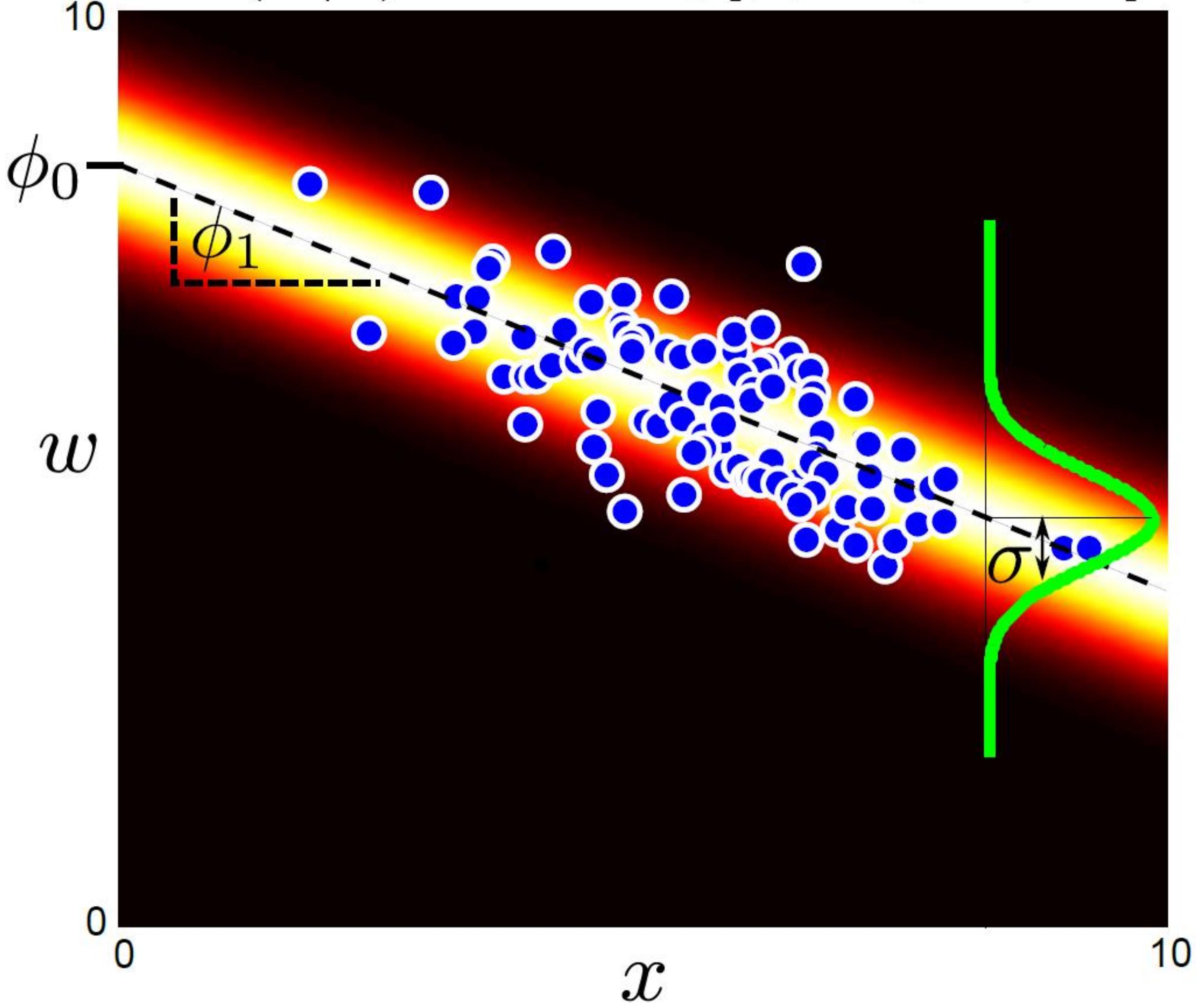
$$Pr(w|x) = \text{Norm}_w[\phi_0 + \phi_1 x, \sigma^2]$$



Learning algorithm: learn θ from training data **x,y**. 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**

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**

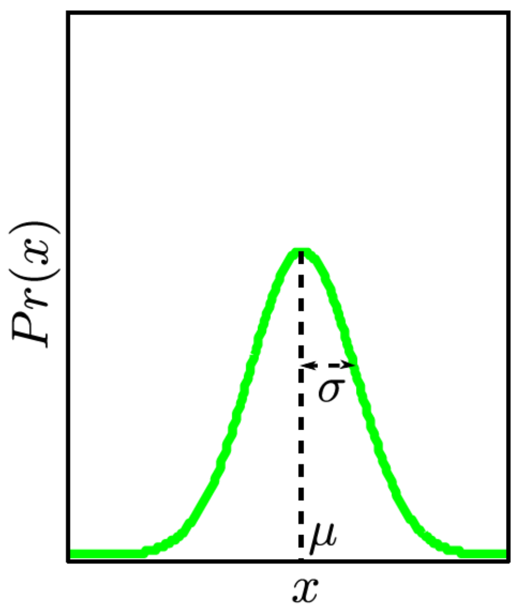
Inference algorithm: Define prior Pr(**w**) and then compute Pr(**w**|**x**) using Bayes' rule

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

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 θ that define its shape

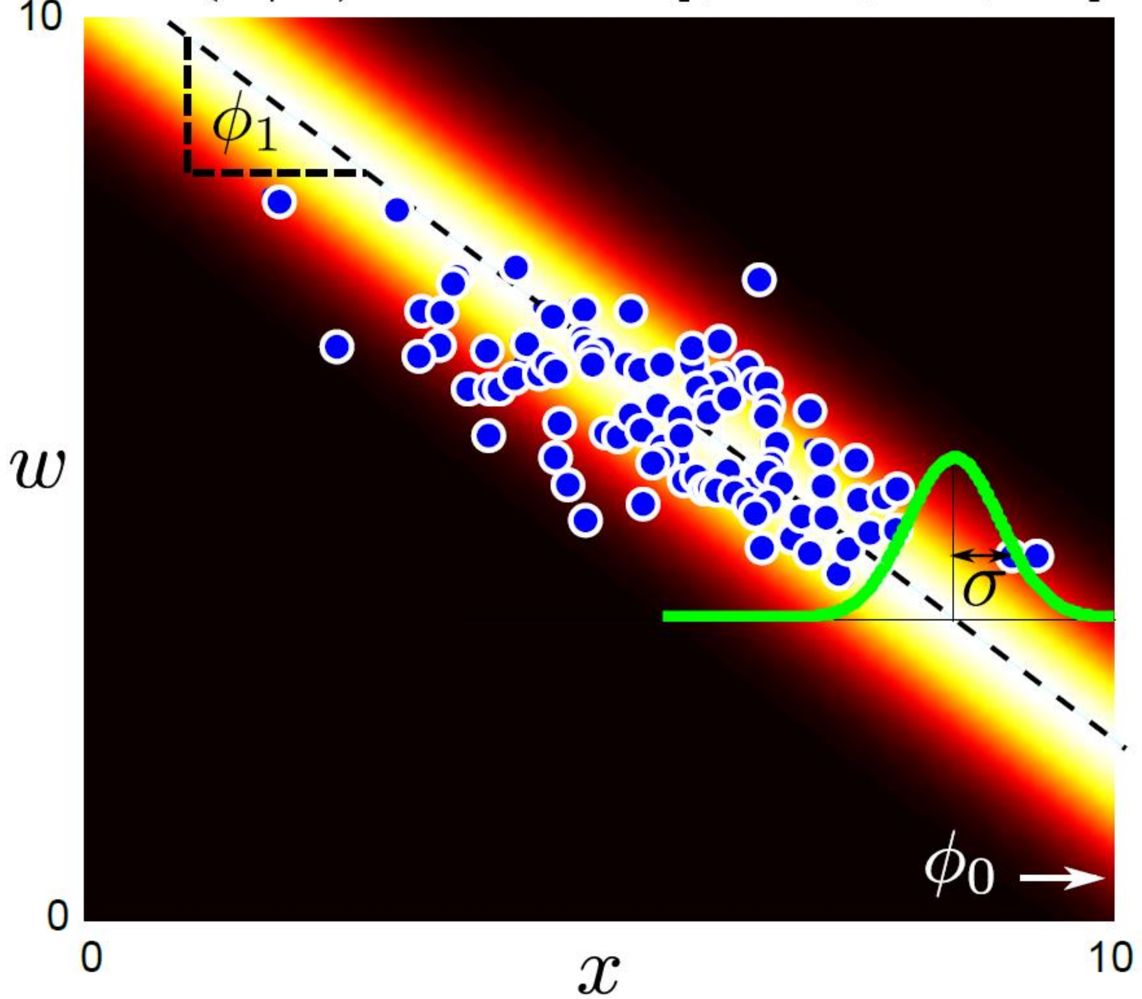


1. Choose normal distribution over x
2. Make mean μ linear function of w (variance constant)

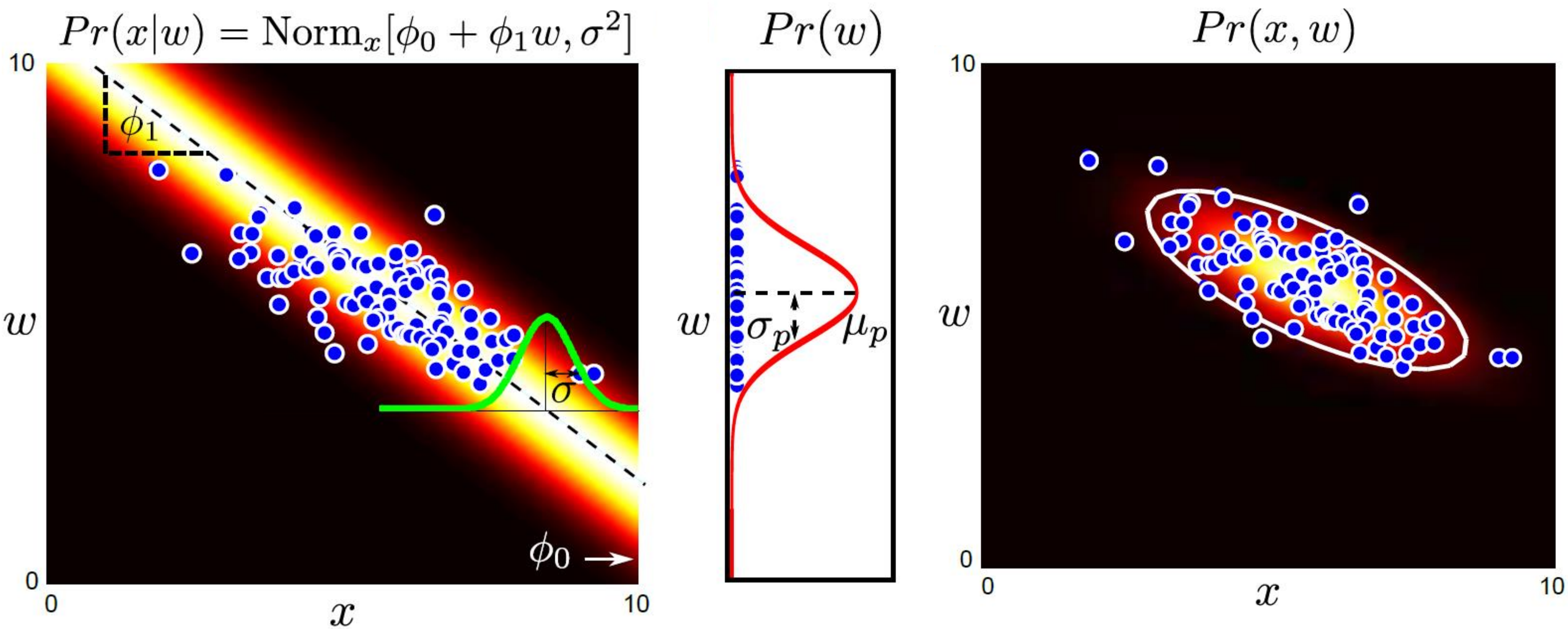
$$Pr(x|w, \boldsymbol{\theta}) = \text{Norm}_x\left[\phi_0 + \phi_1 w, \sigma^2\right]$$

3. Parameter are ϕ_0, ϕ_1, σ^2.

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

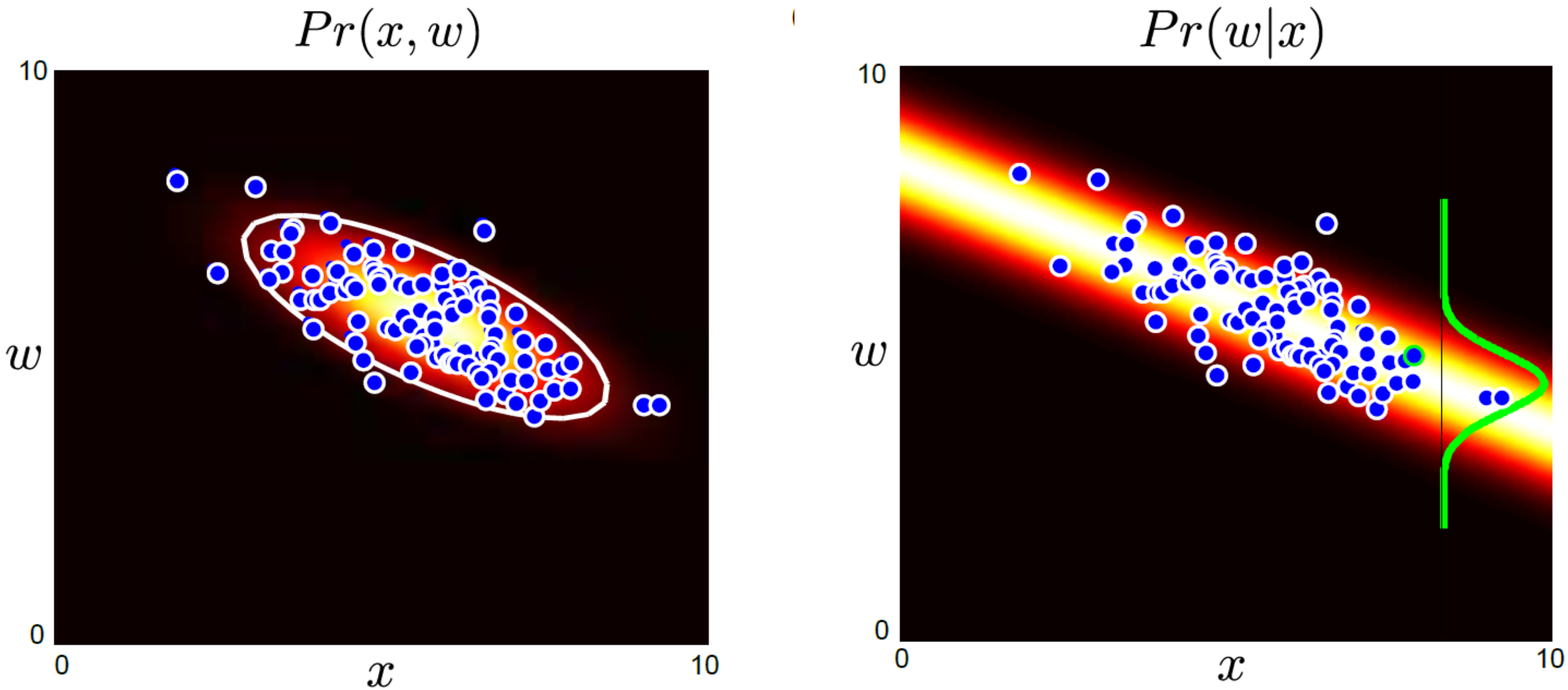


Learning algorithm: learn θ from training data **x**,**w**. e.g. MAP



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

Can get back to joint probability Pr(x,y)



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

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

Structure

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

Worked example 2: Classification

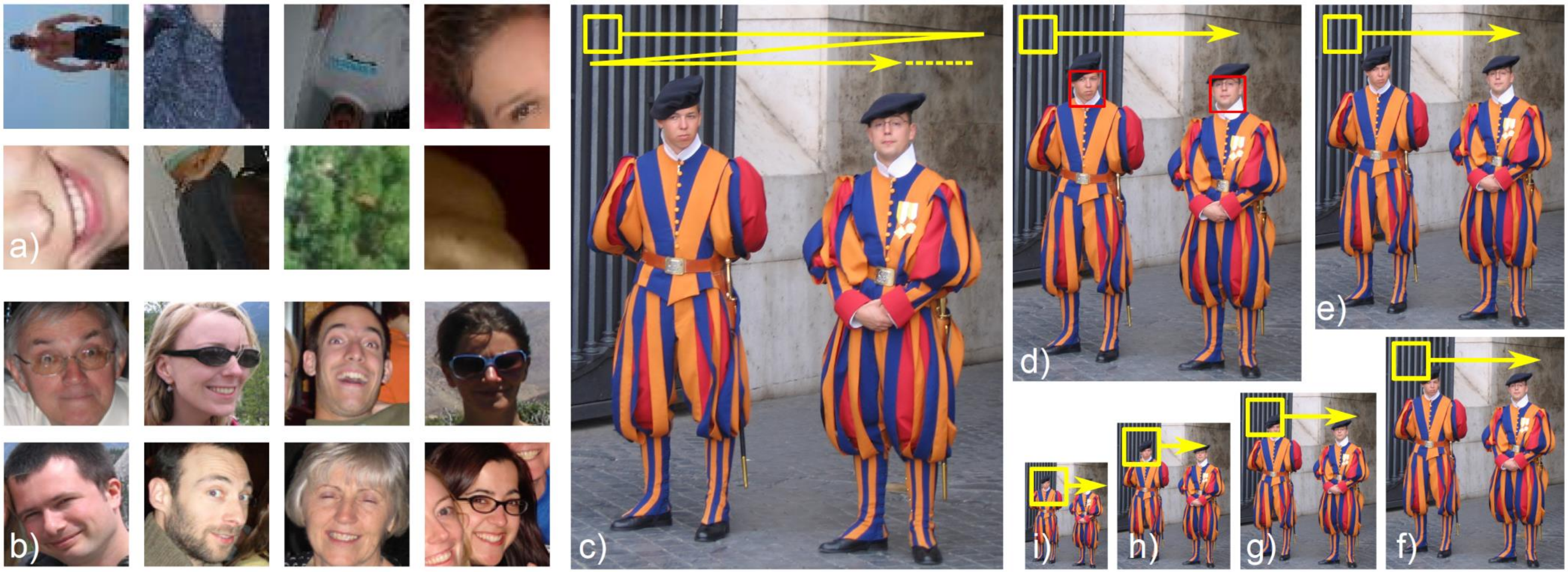
Consider simple case where

- we make a univariate continuous measurement x
- use this to predict a discrete binary world

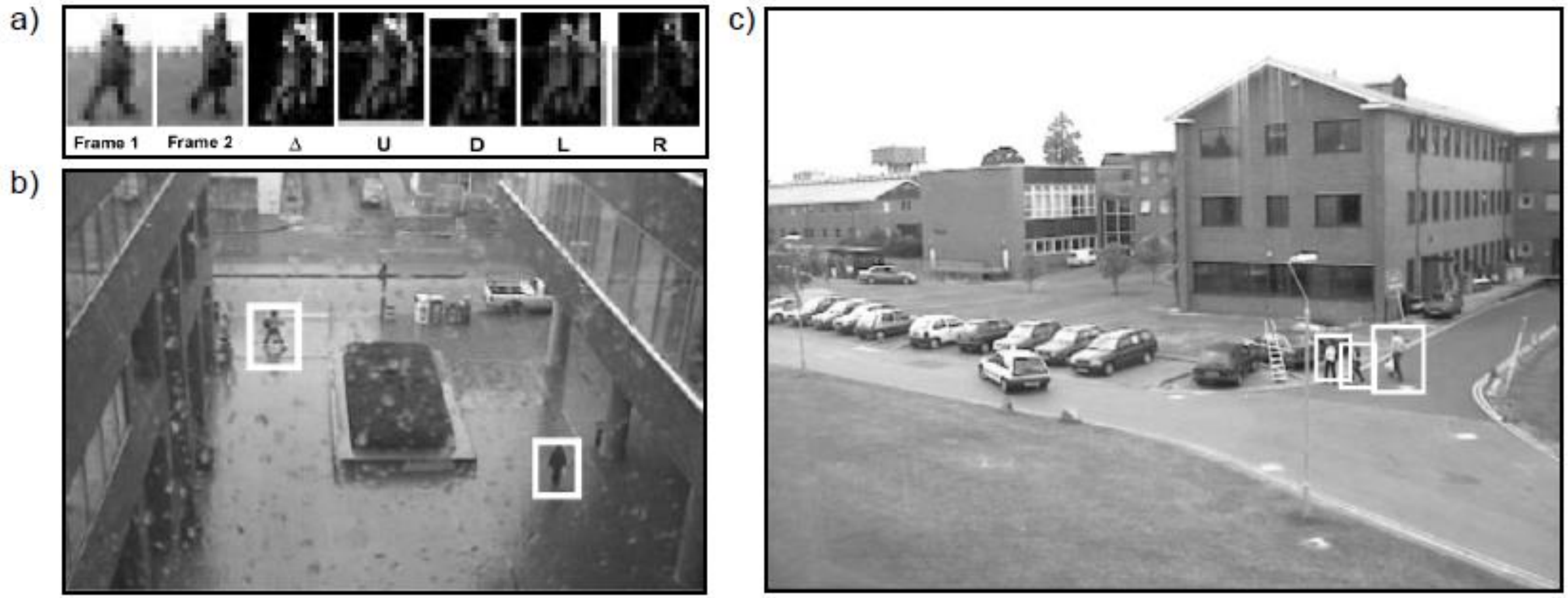
$$w \in \{0, 1\}$$

(classification as world state is discrete)

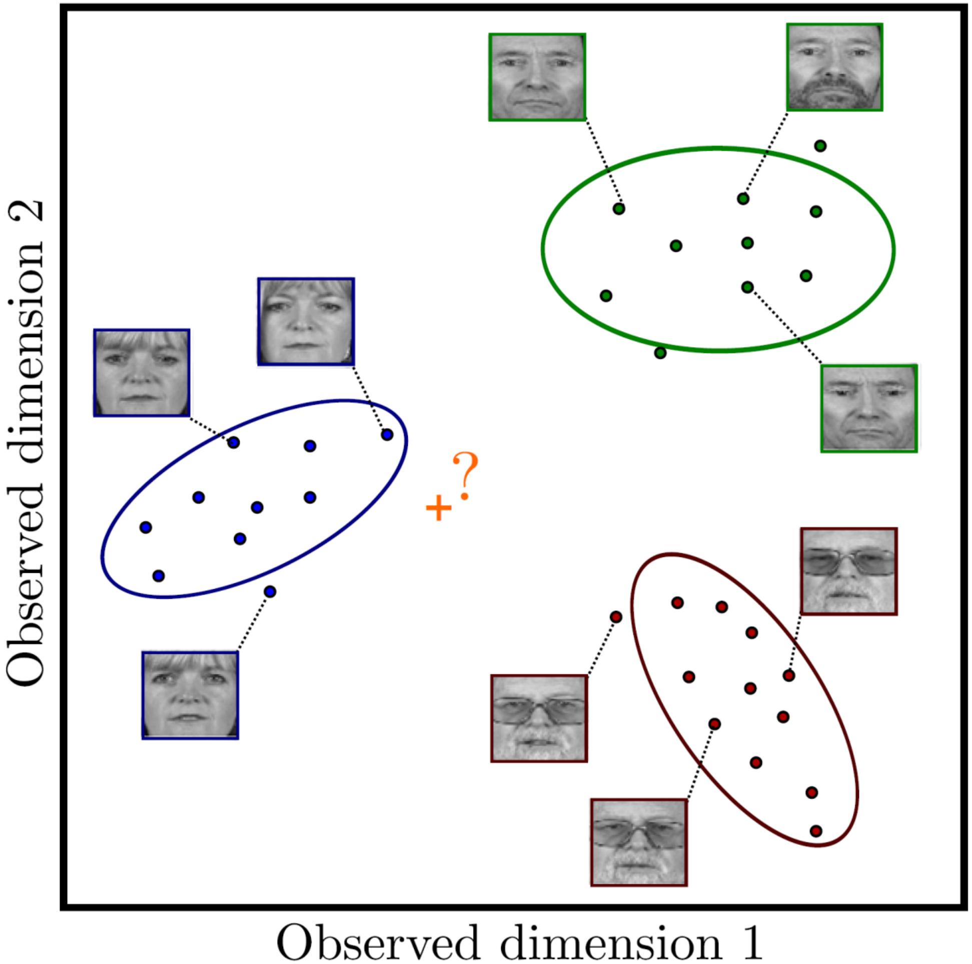
Classification Example 1: Face Detection



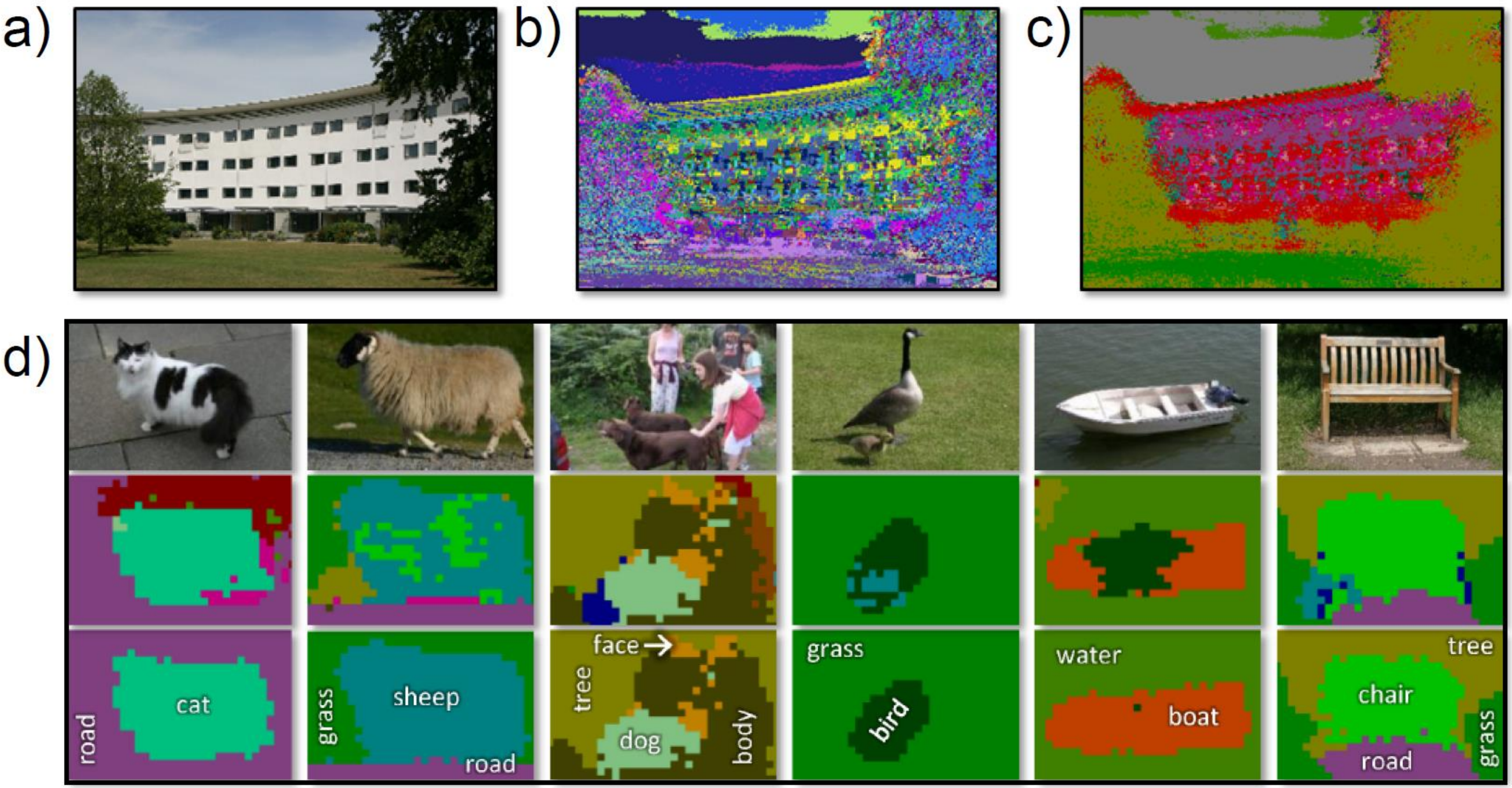
Classification Example 2: Pedestrian Detection



Classification Example 3: Face Recognition



Classification Example 4: Semantic Segmentation



Worked example 2: Classification

Consider simple case where

- we make a univariate continuous measurement x
- use this to predict a discrete binary world

w $\in \{0, 1\}$

(classification as world state is discrete)

Type 1: Model Pr($\mathbf{w}|\mathbf{x}$) - Discriminative

How to model Pr($\mathbf{w}|\mathbf{x}$)?

- Choose an appropriate form for Pr($\mathbf{w}$)
- Make parameters a function of $\mathbf{x}$
- Function takes parameters θ that define its shape

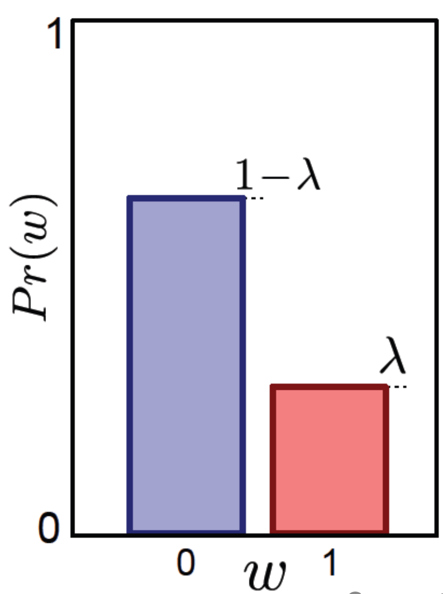
Learning algorithm: learn parameters θ from training data $\mathbf{x},\mathbf{w}$

Inference algorithm: just evaluate **Pr($\mathbf{w}|\mathbf{x}$)**

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

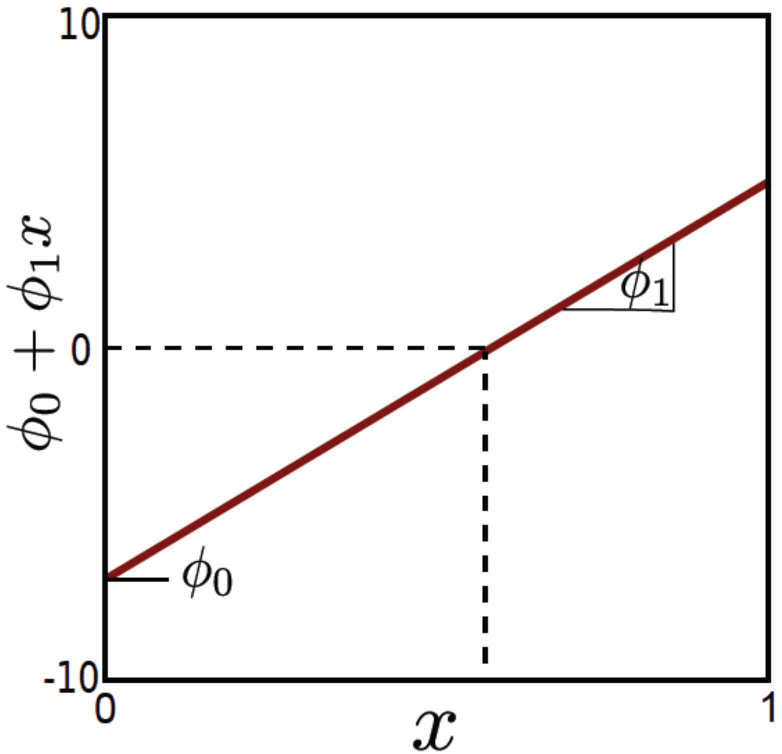


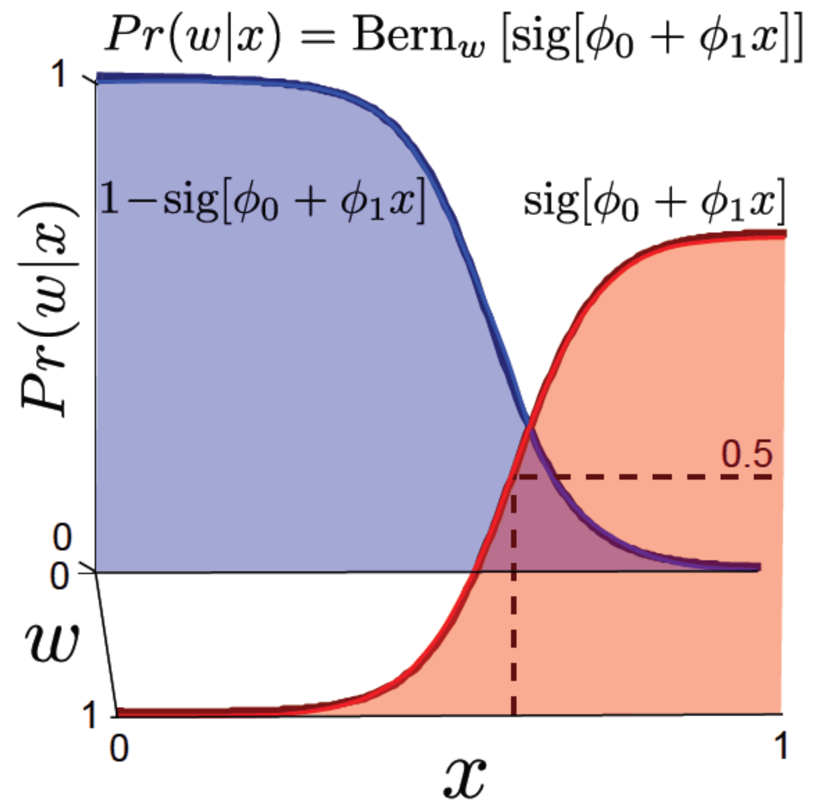
1. Choose Bernoulli dist. for Pr(**w**)
2. Make parameters a function of **x**

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

3. Function takes parameters ϕ_0 and ϕ_1

This model is called *logistic regression.*





Two parameters

$$\boldsymbol{\theta} = \{\phi_0, \phi_1\}$$

Learning by standard methods (ML,MAP, Bayesian)
Inference: 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**

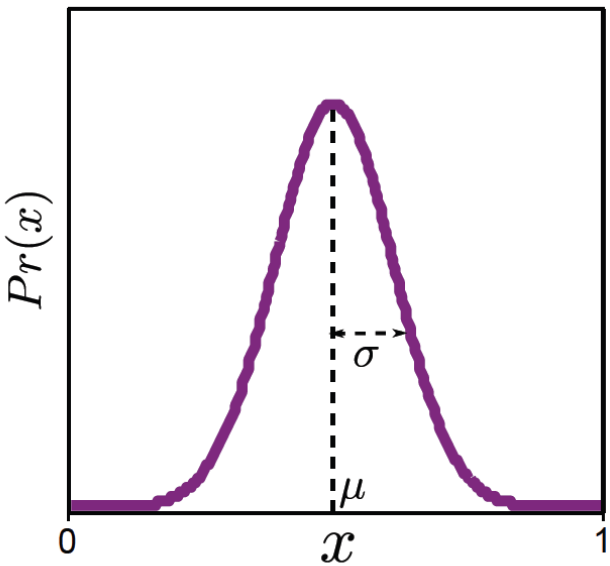
Inference algorithm: Define prior Pr(**w**) and then compute Pr(**w**|**x**) using Bayes' rule

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

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

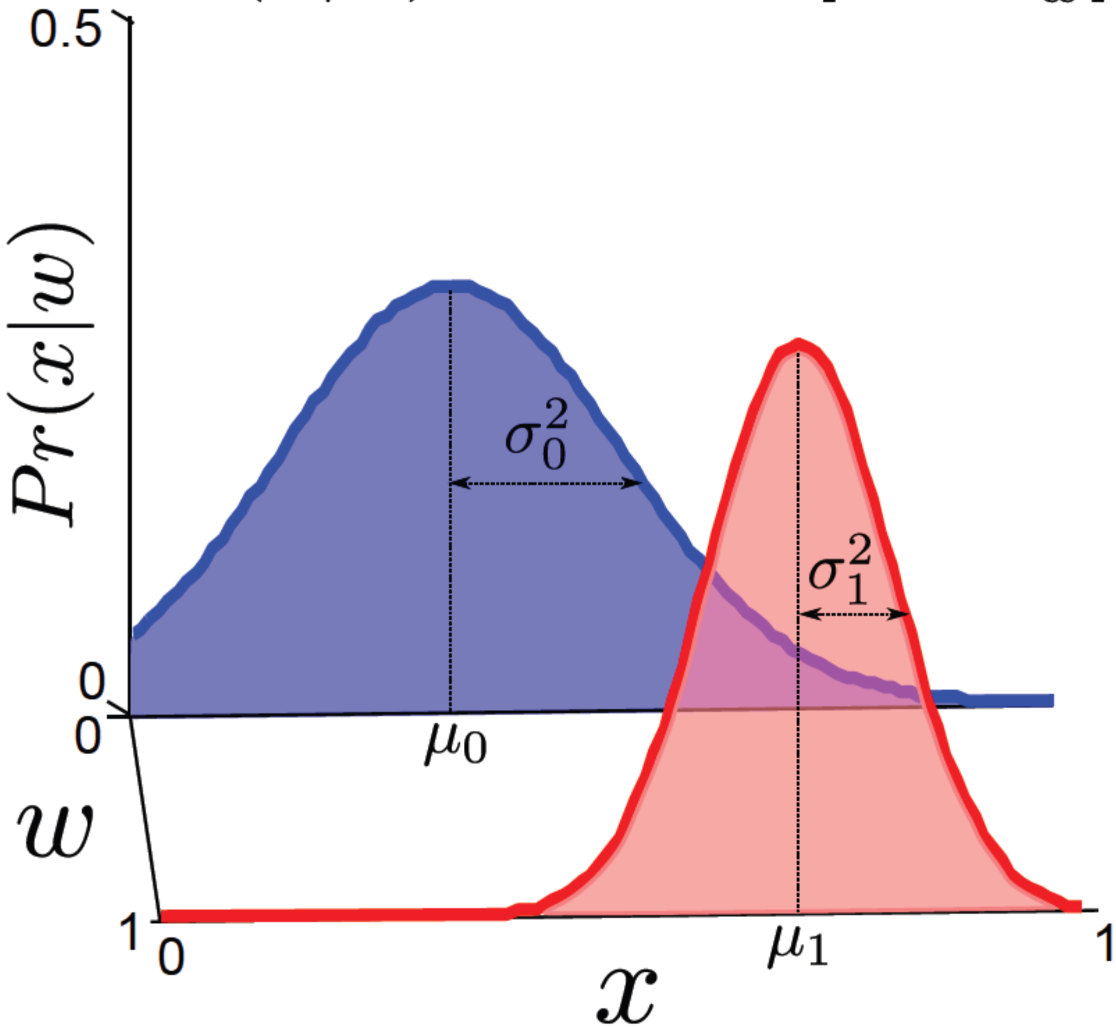


1. Choose a Gaussian distribution for Pr(**x**)
2. Make parameters a function of discrete binary **w**

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

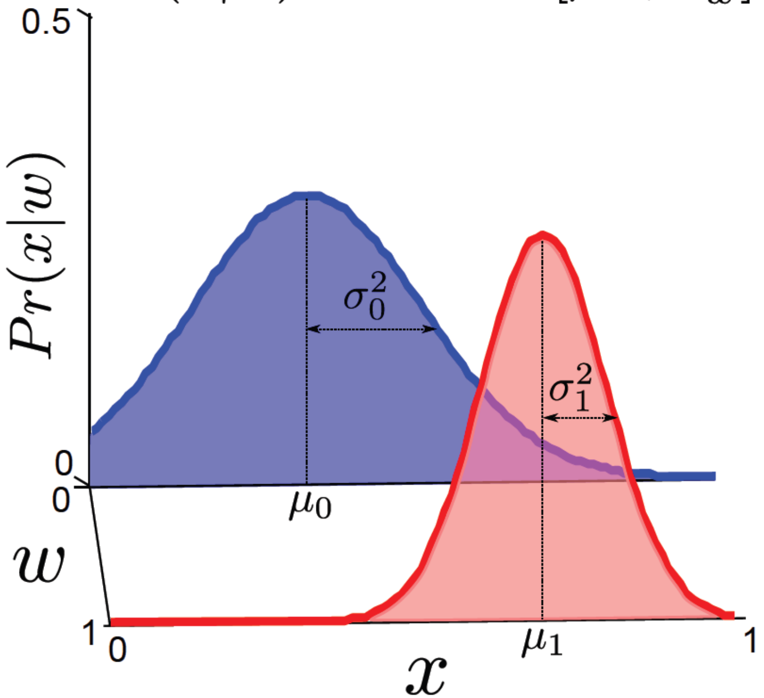
3. Function takes parameters $\boldsymbol{\mu}_0$, $\boldsymbol{\mu}_1$, $\boldsymbol{\sigma}^2{}_0$, $\boldsymbol{\sigma}^2{}_1$ that define its shape

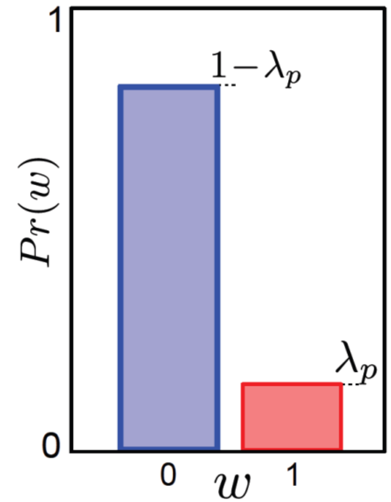
$$Pr(x|w) = \text{Norm}_x[\mu_w, \sigma_w^2]$$

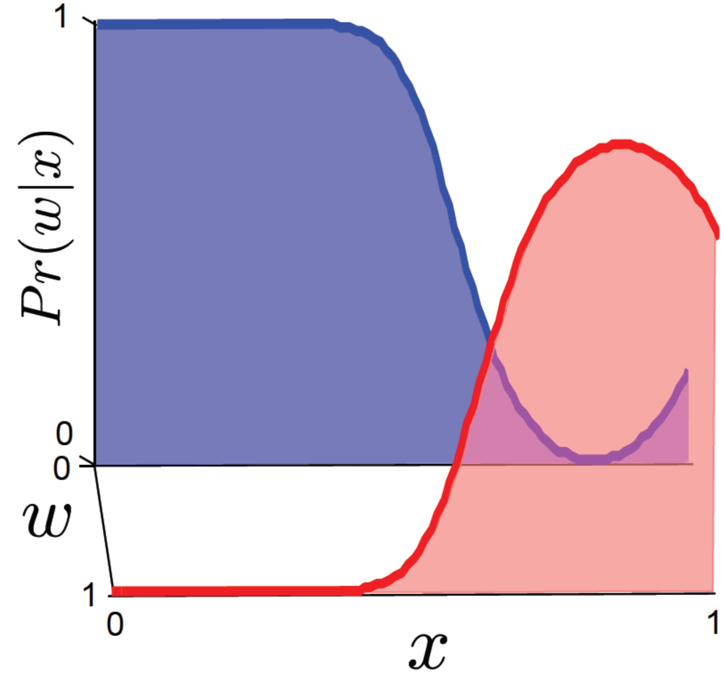


Learn parameters μ_0, μ_1, $\sigma^2{}_0$, $\sigma^2{}_1$ that define its shape

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







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

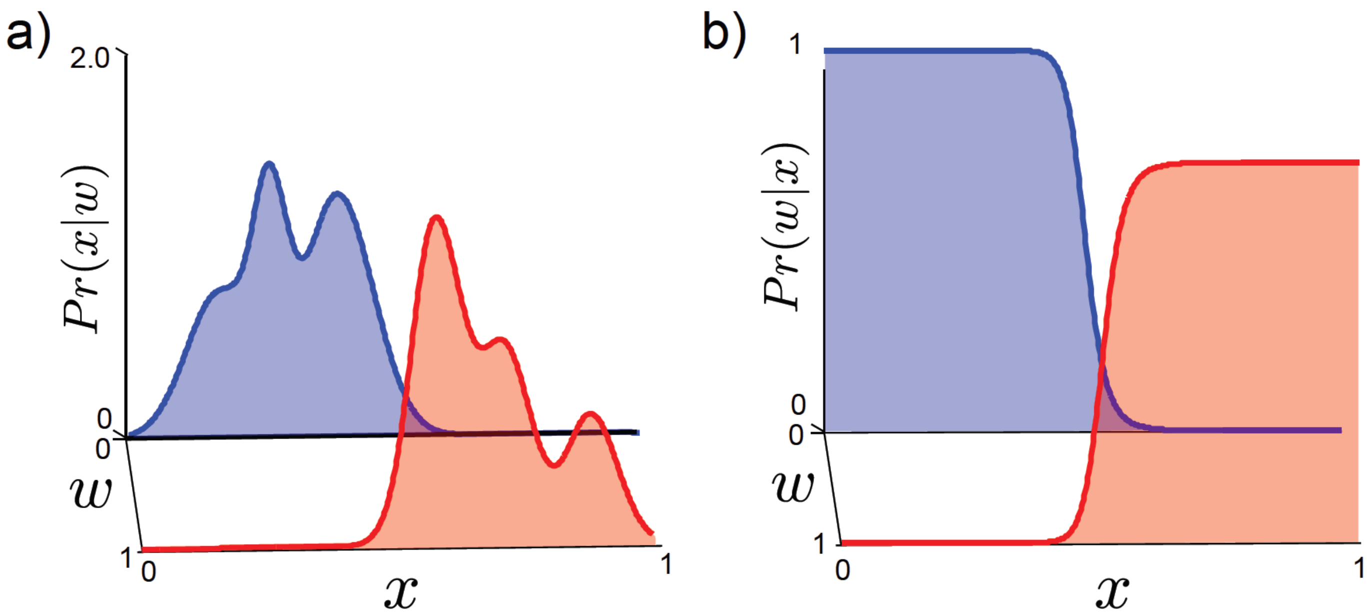
$$Pr(\mathbf{w}|\mathbf{x}) = \frac{Pr(\mathbf{x}|\mathbf{w})Pr(\mathbf{w})}{\int Pr(\mathbf{x}|\mathbf{w})Pr(\mathbf{w})d\mathbf{w}}$$

Structure

- Computer vision models
  - Three types of model
- Worked example 1: Regression
- Worked example 2: Classification
- Which type 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



Which type of model to use?

2. Inference simple in discriminative models
3. Data really is generated from world – generative matches this
4. If missing data, then generative preferred
5. Generative allows imposition of prior knowledge specified by user

Structure

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

Application: Skin Detection

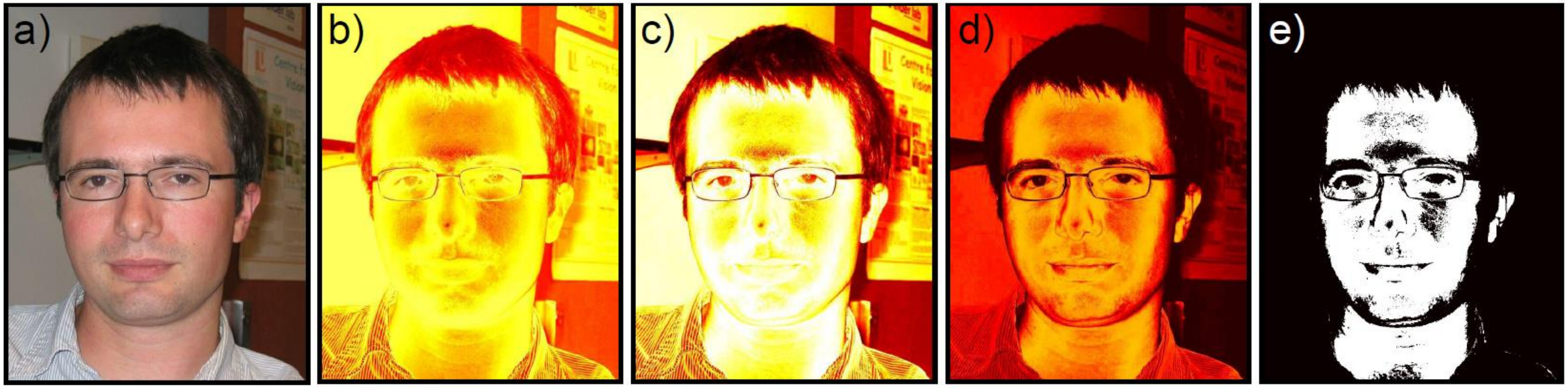
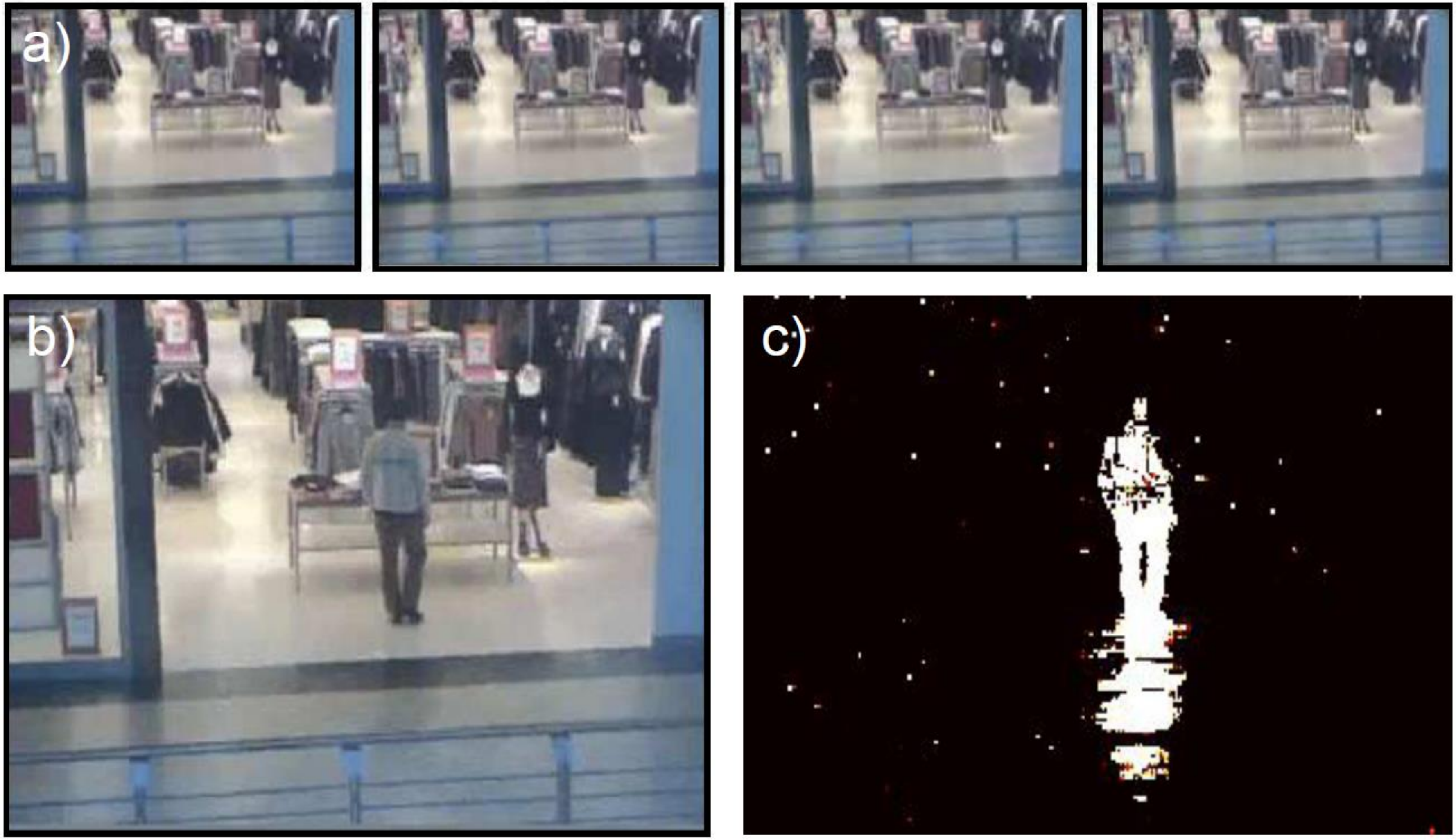
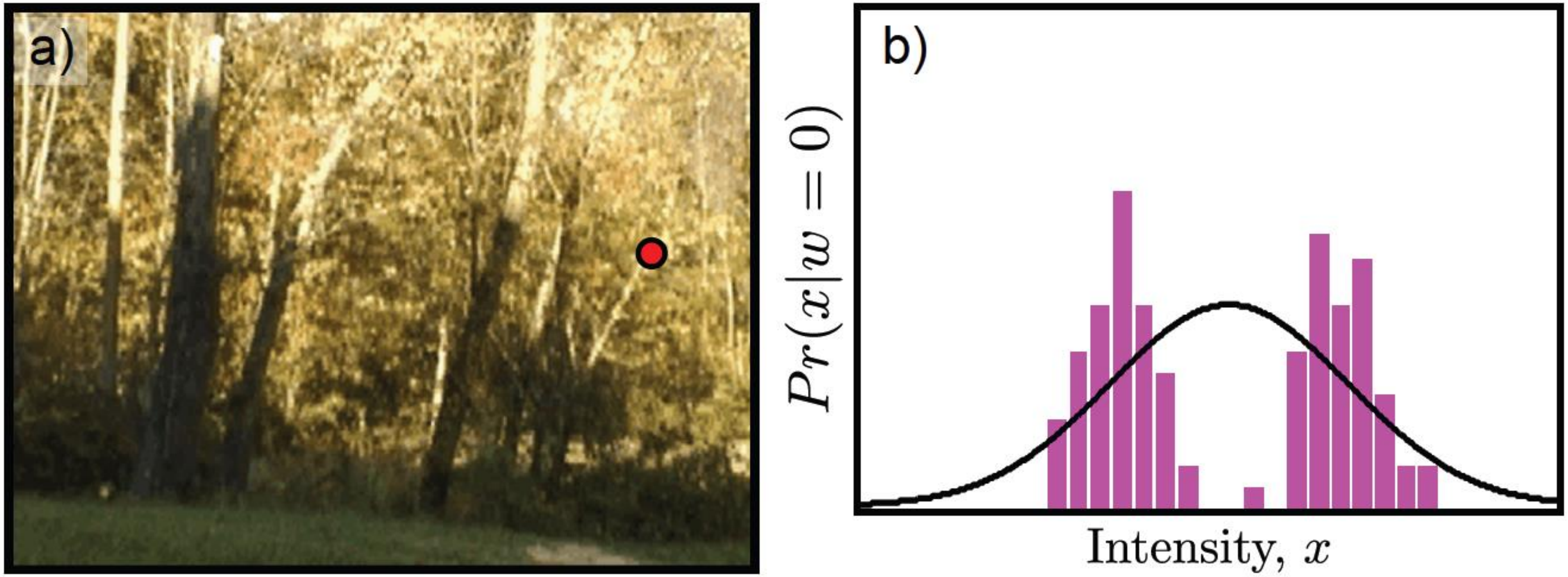


Figure 6.7 Skin detection. For each pixel we aim to infer a label $w \in \{0, 1\}$ denoting the absence or presence of skin based on the RGB triple $\mathbf{x}$. Here we modeled the class conditional density functions $Pr(\mathbf{x}|w)$ as normal distributions. a) Original image. b) Log likelihood (log of data assessed under class-conditional density function) for non-skin. c) Log likelihood for skin. d) Posterior probability of belonging to skin class. e) Thresholded posterior probability $Pr(w|\mathbf{x}) > 0.5$ gives estimate of w.

Application: Background subtraction



Application: Background subtraction



But consider this scene in which the foliage is blowing in the wind. A normal distribution is not good enough! Need a way to make more complex distributions

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

Conclusion

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

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