Building Probabilistic Models

Mark van der Wilk

Department of Computing
Imperial College London

@markvanderwilk
m.vdwilk@imperial.ac.uk

January 16, 2023

Questions

Last time, I gave examples of probabilistic reasoning, e.g. colour perception:

$$P(C|L) = \sum_i P(C|L, I = i)\, P(I = i)$$

$$P(C|L, I = i) = \frac{P(L|C, I = i)P(C)}{P(L|I = i)}$$

Questions

Last time, I gave examples of probabilistic reasoning, e.g. colour perception:

$$P(C|L) = \sum_i P(C|L, I = i)\, P(I = i)$$

$$P(C|L, I = i) = \frac{P(L|C, I = i)P(C)}{P(L|I = i)}$$

- But why this structure?

Questions

Last time, I gave examples of probabilistic reasoning, e.g. colour perception:

$$P(C|L) = \sum_i P(C|L, I = i)\, P(I = i)$$

$$P(C|L, I = i) = \frac{P(L|C, I = i)P(C)}{P(L|I = i)}$$

- But why this structure?
- What assumptions were made?

Questions

Last time, I gave examples of probabilistic reasoning, e.g. colour perception:

$$P(C|L) = \sum_i P(C|L, I = i)\, P(I = i)$$

$$P(C|L, I = i) = \frac{P(L|C, I = i)P(C)}{P(L|I = i)}$$

- But why this structure?
- What assumptions were made?
- How should assumptions be expressed and communicated?

Questions

Last time, I gave examples of probabilistic reasoning, e.g. colour perception:

$$P(C|L) = \sum_i P(C|L, I = i)\, P(I = i)$$

$$P(C|L, I = i) = \frac{P(L|C, I = i)P(C)}{P(L|I = i)}$$

- But why this structure?
- What assumptions were made?
- How should assumptions be expressed and communicated?
- How do we manipulate distributions to the form we want?

Questions

Last time, I gave examples of probabilistic reasoning, e.g. colour perception:

$$P(C|L) = \sum_i P(C|L, I = i)\, P(I = i)$$

$$P(C|L, I = i) = \frac{P(L|C, I = i)P(C)}{P(L|I = i)}$$

- But why this structure?
- What assumptions were made?
- How should assumptions be expressed and communicated?
- How do we manipulate distributions to the form we want?

How to systematically approach
a probabilistic modelling problem?

Mathematical Modelling

Often, we can pose a mathematical model of a phenomenon:

- Reflection (Phong model, different symbol convention)

$$I_{\mathrm{p}} = k_{\mathrm{a}} i_{\mathrm{a}} + \sum_{m \in \text{lights}} \left(k_{\mathrm{d}} (\hat{L}_m \cdot \hat{N}) i_{m,\mathrm{d}} + k_{\mathrm{s}} (\hat{R}_m \cdot \hat{V})^{\alpha} i_{m,\mathrm{s}}\right)$$

Mathematical Modelling

Often, we can pose a mathematical model of a phenomenon:

- Reflection (Phong model, different symbol convention)

$$I_\mathrm{p} = k_\mathrm{a} i_\mathrm{a} + \sum_{m \,\in\, \text{lights}} (k_\mathrm{d} (\hat{L}_m \cdot \hat{N}) i_{m,\mathrm{d}} + k_\mathrm{s} (\hat{R}_m \cdot \hat{V})^\alpha i_{m,\mathrm{s}})$$

- Movement of an object under gravity (Newton's laws)

$$s_t = \frac{1}{2}\frac{F}{m}t^2 + v_0 t \tag{1}$$

Mathematical Modelling

Often, we can pose a mathematical model of a phenomenon:

- Reflection (Phong model, different symbol convention)

$$I_{\mathrm{p}} = k_{\mathrm{a}} i_{\mathrm{a}} + \sum_{m \,\in\, \text{lights}} \left(k_{\mathrm{d}} (\hat{L}_m \cdot \hat{N}) i_{m,\mathrm{d}} + k_{\mathrm{s}} (\hat{R}_m \cdot \hat{V})^{\alpha} i_{m,\mathrm{s}}\right)$$

- Movement of an object under gravity (Newton's laws)

$$s_t = \frac{1}{2}\frac{F}{m}t^2 + v_0 t \tag{1}$$

So if you are **given** some quantities, you can make a prediction about another.

From Mathematical Models to Probabilistic Models

- A mathematical model expresses deterministic relationships.

From Mathematical Models to Probabilistic Models

- A mathematical model expresses deterministic relationships.
- A probabilistic model expresses relationships with uncertainty.

From Mathematical Models to Probabilistic Models

- A mathematical model expresses deterministic relationships.
- A probabilistic model expresses relationships with uncertainty.
- Often, probabilistic models are specified starting with a mathematical model.

From Mathematical Models to Probabilistic Models

- A mathematical model expresses deterministic relationships.
- A probabilistic model expresses relationships with uncertainty.
- Often, probabilistic models are specified starting with a mathematical model.
- Mathematical relationships can help specify conditional distributions.

From Mathematical Models to Probabilistic Models

- A mathematical model expresses deterministic relationships.
- A probabilistic model expresses relationships with uncertainty.
- Often, probabilistic models are specified starting with a mathematical model.
- Mathematical relationships can help specify conditional distributions.

Mathematical models are a *special case* of deterministic models.
Probability can still express certainty!

From Mathematical Models to Probabilistic Models

- A mathematical model expresses deterministic relationships.
- A probabilistic model expresses relationships with uncertainty.
- Often, probabilistic models are specified starting with a mathematical model.
- Mathematical relationships can help specify conditional distributions.

Mathematical models are a *special case* of deterministic models.
Probability can still express certainty!
E.g. from Newton's laws:

$$s_t = \frac{1}{2}\frac{F}{m}t^2 + v_0 t \tag{2}$$

$$p(s_t|v_0, m, F) = \delta(s_t - \frac{1}{2}\frac{F}{m}t^2 + v_0 t) \tag{3}$$

(Remember: $\int_{\mathcal{R}} \delta(\mathbf{x} - \mathbf{y}) \mathrm{d}\mathbf{x} = 1$ if $\mathbf{y} \in \mathcal{R}$, 0 otherwise.)

Probabilistic Models: Uncertain Quantities

Given a mathematical model.

$$p(s_t|v_0, m, F) = \delta(s_t - \frac{1}{2}\frac{F}{m}t^2 + v_0 t) \tag{4}$$

A certain relationship like this can be used to work back to uncertain quantities.

Probabilistic Models: Uncertain Quantities

Given a mathematical model.

$$p(s_t|v_0, m, F) = \delta(s_t - \frac{1}{2}\frac{F}{m}t^2 + v_0 t) \tag{4}$$

A certain relationship like this can be used to work back to uncertain quantities. Imagine we are uncertain about certain quantities:

$$p(v_0, m, F) = \mathcal{N}(v_0; \mu_v, 1.0)\delta(m - 1.0)\mathcal{N}(F; \mu_F, 0.1) \tag{5}$$

Probabilistic Models: Uncertain Quantities

Given a mathematical model.

$$p(s_t|v_0, m, F) = \delta(s_t - \frac{1}{2}\frac{F}{m}t^2 + v_0 t) \tag{4}$$

A certain relationship like this can be used to work back to uncertain quantities. Imagine we are uncertain about certain quantities:

$$p(v_0, m, F) = \mathcal{N}(v_0; \mu_v, 1.0)\delta(m - 1.0)\mathcal{N}(F; \mu_F, 0.1) \tag{5}$$

We can find how our uncertainty over the initial velocity v_0 changes by finding $p(v_0|s_t)$!

Probabilistic Models: Uncertain Relationships

- Is the relationship $p(s_t|v_0, m, F) = \delta(s_t - \frac{1}{2}\frac{F}{m}t^2 + v_0 t)$ **realistic**?

Probabilistic Models: Uncertain Relationships

- Is the relationship $p(s_t|v_0, m, F) = \delta(s_t - \frac{1}{2}\frac{F}{m}t^2 + v_0 t)$ **realistic**?
- If we did an initial value experiment, would we really measure **exactly** the predicted value?

Probabilistic Models: Uncertain Relationships

- Is the relationship $p(s_t|v_0, m, F) = \delta(s_t - \frac{1}{2}\frac{F}{m}t^2 + v_0 t)$ **realistic**?
- If we did an initial value experiment, would we really measure **exactly** the predicted value?
- Adding uncertainty makes predictions more **realistic**, by allowing **errors**.

$$p(s_t|v_0, m, F) = \mathcal{N}\left(s_t; \frac{1}{2}\frac{F}{m}t^2 + v_0 t, \sigma^2\right) \tag{6}$$

Probability of Everything

How to make these intuitions **systematic**?

Probability of Everything

How to make these intuitions **systematic**?

It's a good idea to start from a representation that:

1. clearly expresses the assumptions made in our model,
2. allows us to derive any distribution that we are interested in.

Probability of Everything

How to make these intuitions **systematic**?
It's a good idea to start from a representation that:

1. clearly expresses the assumptions made in our model,
2. allows us to derive any distribution that we are interested in.

The joint distribution over all variables.
The **Probability of Everything**.

Probability of Everything

How to make these intuitions **systematic**?
It's a good idea to start from a representation that:

1. clearly expresses the assumptions made in our model,
2. allows us to derive any distribution that we are interested in.

The joint distribution over all variables.
The **Probability of Everything**.

$$p(x, y, z) \tag{7}$$

Probability of Everything

How to make these intuitions **systematic**?
It's a good idea to start from a representation that:

1. clearly expresses the assumptions made in our model,
2. allows us to derive any distribution that we are interested in.

The joint distribution over all variables.
The **Probability of Everything**.

$$p(x, y, z) \tag{7}$$

Any question we may want to answer corresponds to finding:

$$p(x, z|y) = \frac{p(x, y, z)}{p(y)} \qquad \text{or} \qquad p(x|y) = \int \frac{p(x, y, z)}{p(y)} \mathrm{d}z \tag{8}$$

Probability of Everything

How to make these intuitions **systematic**?
It's a good idea to start from a representation that:

1. clearly expresses the assumptions made in our model,
2. allows us to derive any distribution that we are interested in.

The joint distribution over all variables.
The **Probability of Everything**.

$$p(x, y, z) \tag{7}$$

Any question we may want to answer corresponds to finding:

$$p(x, z|y) = \frac{p(x, y, z)}{p(y)} \qquad \text{or} \qquad p(x|y) = \int \frac{p(x, y, z)}{p(y)} \mathrm{d}z \tag{8}$$

- Observe a variable? Conditioning (i.e. divide and renormalise).
- Not interested in a variable? Marginalise.

Building a Probabilistic Model: Statistical Approach

Understanding how your variables causally interact gives you a factorisation of the joint.

Building a Probabilistic Model: Statistical Approach

Understanding how your variables causally interact gives you a factorisation of the joint.

1. Identify all the variables that are relevant to your problem.

Building a Probabilistic Model: Statistical Approach

Understanding how your variables causally interact gives you a factorisation of the joint.

1. Identify all the variables that are relevant to your problem.
2. Start with a variable that you will observe.

Building a Probabilistic Model: Statistical Approach

Understanding how your variables causally interact gives you a factorisation of the joint.

1. Identify all the variables that are relevant to your problem.
2. Start with a variable that you will observe.
3. Determine which variables it depends on, and choose a sensible conditional distribution.

Building a Probabilistic Model: Statistical Approach

Understanding how your variables causally interact gives you a factorisation of the joint.

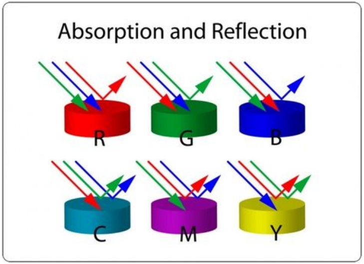
1. Identify all the variables that are relevant to your problem.
2. Start with a variable that you will observe.
3. Determine which variables it depends on, and choose a sensible conditional distribution.
4. Repeat previous step for one of the variables that you conditioned on.

Example: Lighting

Step 1: Identify all variables:

- Object colour C.
- Reflected light L.
- Illumination I.

Joint: $p(C, L, I)$.

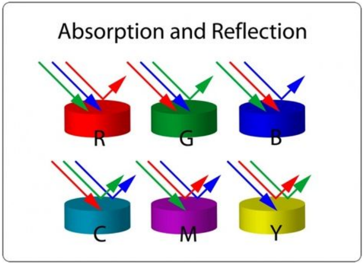


Example: Lighting

Step 1: Identify all variables:

- Object colour C.
- Reflected light L.
- Illumination I.

Joint: $p(C, L, I)$.



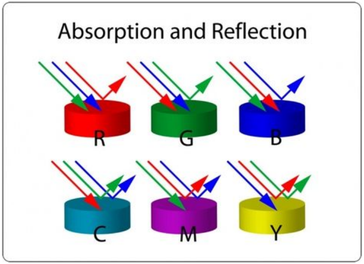
Step 2: We observe the reflected light L. We have a model for the L given I and C: $p(L, C, I) = p(L|C, I)p(C, I)$.

Example: Lighting

Step 1: Identify all variables:

- Object colour C.
- Reflected light L.
- Illumination I.

Joint: $p(C, L, I)$.



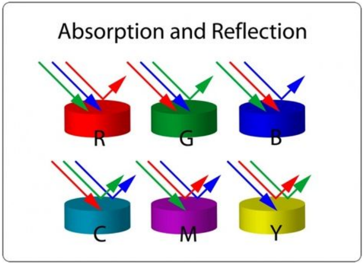
Step 2: We observe the reflected light L. We have a model for the L given I and C: $p(L, C, I) = p(L|C, I)p(C, I)$.
Step 3: Let's pick C. Colour does not depend on illumination, so $p(C|I) = p(C)$, showing that C and I are independent.

Example: Lighting

Step 1: Identify all variables:

- Object colour C.
- Reflected light L.
- Illumination I.

Joint: $p(C, L, I)$.



Step 2: We observe the reflected light L. We have a model for the L given I and C: $p(L, C, I) = p(L|C, I)p(C, I)$.
Step 3: Let's pick C. Colour does not depend on illumination, so $p(C|I) = p(C)$, showing that C and I are independent.

While we can use our knowledge for choosing $p(L|C, I)$, we need to choose subjective priors for $p(C)$ and $p(I)$.

Finding the right posterior

- Now that we have the joint, how do we find $p(C|L)$?
- Remember: We need to find it in terms of the conditional distributions **which we can actually evaluate**.
- This is why starting with the joint is such a good idea! Given the definitions from the previous slide, we can evaluate it!

$$p(C|L) = \frac{p(C, L)}{p(L)} = \tag{9}$$

$$\tag{10}$$

Finding the right posterior

- Now that we have the joint, how do we find $p(C|L)$?
- Remember: We need to find it in terms of the conditional distributions **which we can actually evaluate**.
- This is why starting with the joint is such a good idea! Given the definitions from the previous slide, we can evaluate it!

$$p(C|L) = \frac{p(C,L)}{p(L)} = \frac{\sum_I p(C,L,I)}{\sum_{C,I} p(C,L,I)} \tag{9}$$

$$\tag{10}$$

Finding the right posterior

- Now that we have the joint, how do we find $p(C|L)$?
- Remember: We need to find it in terms of the conditional distributions **which we can actually evaluate**.
- This is why starting with the joint is such a good idea! Given the definitions from the previous slide, we can evaluate it!

$$p(C|L) = \frac{p(C, L)}{p(L)} = \frac{\sum_I p(C, L, I)}{\sum_{C,I} p(C, L, I)} \tag{9}$$

$$= \frac{\sum_I p(L|C, I)\, p(C)\, p(I)}{\dots} \tag{10}$$

Finding the right posterior

- Now that we have the joint, how do we find $p(C|L)$?
- Remember: We need to find it in terms of the conditional distributions **which we can actually evaluate**.
- This is why starting with the joint is such a good idea! Given the definitions from the previous slide, we can evaluate it!

$$p(C|L) = \frac{p(C, L)}{p(L)} = \frac{\sum_I p(C, L, I)}{\sum_{C,I} p(C, L, I)} \tag{9}$$

$$= \frac{\sum_I p(L|C, I)\, p(C)\, p(I)}{\ldots} \tag{10}$$

We can take many different routes!

$$p(C|L) = \frac{p(L|C)\, p(C)}{p(L)} = \frac{\left[\sum_I p(L|C, I) p(I)\right] p(C)}{p(L)} \tag{11}$$

Many roads lead to Rome, but starting from the joint highlights assumptions

Example: Burglars, Earthquakes, and Alarms

"Fred lives in Los Angeles and commutes 60 miles to work. Whilst at work, he receives a phone-call from his neighbour saying that Fred's burglar alarm is ringing. What is the probability that there was a burglar in his house today? While driving home to investigate, Fred hears on the radio that there was a small earthquake that day near his home. 'Oh', he says, feeling relieved, 'it was probably the earthquake that set off the alarm'. What is the probability that there was a burglar in his house?" (MacKay, 2003, §21.1)

Q: How does the joint factorise? What conditionals should we define?

- Variables: **p**honecall, **a**larm, **b**urglar, **r**adio, **e**arthquake

$$p(p, a, b, r, e) = p(p, a, b, r, e) \tag{12}$$

Example: Burglars, Earthquakes, and Alarms

"Fred lives in Los Angeles and commutes 60 miles to work. Whilst at work, he receives a phone-call from his neighbour saying that Fred's burglar alarm is ringing. What is the probability that there was a burglar in his house today? While driving home to investigate, Fred hears on the radio that there was a small earthquake that day near his home. 'Oh', he says, feeling relieved, 'it was probably the earthquake that set off the alarm'. What is the probability that there was a burglar in his house?" (MacKay, 2003, §21.1)

Q: How does the joint factorise? What conditionals should we define?

- Variables: **p**honecall, **a**larm, **b**urglar, **r**adio, **e**arthquake

$$p(p,a,b,r,e) = p(p|a,b,r,e)p(a,b,r,e) \tag{12}$$

Example: Burglars, Earthquakes, and Alarms

"Fred lives in Los Angeles and commutes 60 miles to work. Whilst at work, he receives a phone-call from his neighbour saying that Fred's burglar alarm is ringing. What is the probability that there was a burglar in his house today? While driving home to investigate, Fred hears on the radio that there was a small earthquake that day near his home. 'Oh', he says, feeling relieved, 'it was probably the earthquake that set off the alarm'. What is the probability that there was a burglar in his house?" (MacKay, 2003, §21.1)

Q: How does the joint factorise? What conditionals should we define?

- Variables: **p**honecall, **a**larm, **b**urglar, **r**adio, **e**arthquake

$$p(p, a, b, r, e) = p(p|a)p(a, b, r, e) \tag{12}$$

Example: Burglars, Earthquakes, and Alarms

"Fred lives in Los Angeles and commutes 60 miles to work. Whilst at work, he receives a phone-call from his neighbour saying that Fred's burglar alarm is ringing. What is the probability that there was a burglar in his house today? While driving home to investigate, Fred hears on the radio that there was a small earthquake that day near his home. 'Oh', he says, feeling relieved, 'it was probably the earthquake that set off the alarm'. What is the probability that there was a burglar in his house?" (MacKay, 2003, §21.1)

Q: How does the joint factorise? What conditionals should we define?

- Variables: **p**honecall, **a**larm, **b**urglar, **r**adio, **e**arthquake

$$p(p, a, b, r, e) = p(p|a)p(a|b, r, e)p(b, r, e) \tag{12}$$

Example: Burglars, Earthquakes, and Alarms

"Fred lives in Los Angeles and commutes 60 miles to work. Whilst at work, he receives a phone-call from his neighbour saying that Fred's burglar alarm is ringing. What is the probability that there was a burglar in his house today? While driving home to investigate, Fred hears on the radio that there was a small earthquake that day near his home. 'Oh', he says, feeling relieved, 'it was probably the earthquake that set off the alarm'. What is the probability that there was a burglar in his house?" (MacKay, 2003, §21.1)

Q: How does the joint factorise? What conditionals should we define?

- Variables: **p**honecall, **a**larm, **b**urglar, **r**adio, **e**arthquake

$$p(p, a, b, r, e) = p(p|a)p(a|b, e)p(b, r, e) \tag{12}$$

Example: Burglars, Earthquakes, and Alarms

"Fred lives in Los Angeles and commutes 60 miles to work. Whilst at work, he receives a phone-call from his neighbour saying that Fred's burglar alarm is ringing. What is the probability that there was a burglar in his house today? While driving home to investigate, Fred hears on the radio that there was a small earthquake that day near his home. 'Oh', he says, feeling relieved, 'it was probably the earthquake that set off the alarm'. What is the probability that there was a burglar in his house?" (MacKay, 2003, §21.1)

Q: How does the joint factorise? What conditionals should we define?

- Variables: **p**honecall, **a**larm, **b**urglar, **r**adio, **e**arthquake

$$p(p, a, b, r, e) = p(p|a)p(a|b, e)p(b|r, e)p(r, e) \tag{12}$$

Example: Burglars, Earthquakes, and Alarms

"Fred lives in Los Angeles and commutes 60 miles to work. Whilst at work, he receives a phone-call from his neighbour saying that Fred's burglar alarm is ringing. What is the probability that there was a burglar in his house today? While driving home to investigate, Fred hears on the radio that there was a small earthquake that day near his home. 'Oh', he says, feeling relieved, 'it was probably the earthquake that set off the alarm'. What is the probability that there was a burglar in his house?" (MacKay, 2003, §21.1)

Q: How does the joint factorise? What conditionals should we define?

- Variables: **p**honecall, **a**larm, **b**urglar, **r**adio, **e**arthquake

$$p(p, a, b, r, e) = p(p|a)p(a|b, e)p(b)p(r, e) \tag{12}$$

Example: Burglars, Earthquakes, and Alarms

"Fred lives in Los Angeles and commutes 60 miles to work. Whilst at work, he receives a phone-call from his neighbour saying that Fred's burglar alarm is ringing. What is the probability that there was a burglar in his house today? While driving home to investigate, Fred hears on the radio that there was a small earthquake that day near his home. 'Oh', he says, feeling relieved, 'it was probably the earthquake that set off the alarm'. What is the probability that there was a burglar in his house?" (MacKay, 2003, §21.1)

Q: How does the joint factorise? What conditionals should we define?

- Variables: **p**honecall, **a**larm, **b**urglar, **r**adio, **e**arthquake

$$p(p, a, b, r, e) = p(p|a)p(a|b, e)p(b)p(r|e)p(e) \tag{12}$$

Example: Burglars, Earthquakes, and Alarms

"Fred lives in Los Angeles and commutes 60 miles to work. Whilst at work, he receives a phone-call from his neighbour saying that Fred's burglar alarm is ringing. What is the probability that there was a burglar in his house today? While driving home to investigate, Fred hears on the radio that there was a small earthquake that day near his home. 'Oh', he says, feeling relieved, 'it was probably the earthquake that set off the alarm'. What is the probability that there was a burglar in his house?" (MacKay, 2003, §21.1)

Q: How does the joint factorise? What conditionals should we define?

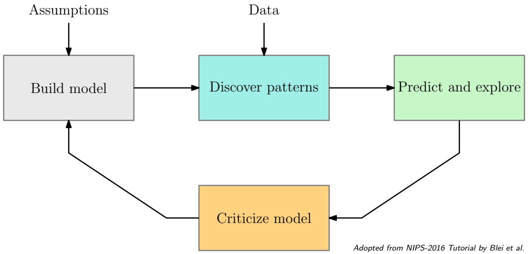
- Variables: **p**honecall, **a**larm, **b**urglar, **r**adio, **e**arthquake

$$p(p, a, b, r, e) = p(p|a)p(a|b, e)p(b)p(r|e)p(e) \tag{12}$$

- After selecting sensible conditionals, we have the joint.

Probabilistic Pipeline

If your assumptions are good/correct, inference will give accurate results and good predictions.



Adopted from NIPS-2016 Tutorial by Blei et al.

- Good models generate data that is similar to the data we observe.
- Predict and explore: Sample from the prior, assess predictions.
- **Criticize/revise the model**.

Building a Probabilistic Model: ML Approach

Q: What happens if we don't have a mathematical/mechanistic model?

- For some problems, little is known about the process.
- No known latent variables to use for creating a model.
- We mainly want good prediction!

Building a Probabilistic Model: ML Approach

Q: What happens if we don't have a mathematical/mechanistic model?

- For some problems, little is known about the process.
- No known latent variables to use for creating a model.
- We mainly want good prediction!

All we really need is $p(\mathcal{D}_{\text{future}}, \mathcal{D}_{\text{observed}})$, so we can find the predictive distribution:

$$p(\mathcal{D}_{\text{future}}|\mathcal{D}_{\text{observed}}) = \frac{p(\mathcal{D}_{\text{future}}, \mathcal{D}_{\text{observed}})}{p(\mathcal{D}_{\text{observed}})} \,. \tag{13}$$

Latent Variables

How do we create a joint with interesting relationships between the observed and future data?

Latent Variables

How do we create a joint with interesting relationships between the observed and future data?

- Invent **latent variables** that are **common** to all data.

$$p(\mathcal{D}_{\text{fut}}, \mathcal{D}_{\text{obs}}, \mathbf{z}) \tag{14}$$

Latent Variables

How do we create a joint with interesting relationships between the observed and future data?

- Invent **latent variables** that are **common** to all data.

$$p(\mathcal{D}_{\text{fut}}, \mathcal{D}_{\text{obs}}, \mathbf{z}) \tag{14}$$

- For simplicity, often data is iid given latent variables.

$$p(\mathcal{D}_{\text{fut}}, \mathcal{D}_{\text{obs}}, \mathbf{z}) = \prod_i p(\mathcal{D}_i | \mathbf{z}) p(\mathbf{z}) \tag{15}$$

Latent Variables

How do we create a joint with interesting relationships between the observed and future data?

- Invent **latent variables** that are **common** to all data.

$$p(\mathcal{D}_{\text{fut}}, \mathcal{D}_{\text{obs}}, \mathbf{z}) \tag{14}$$

- For simplicity, often data is iid given latent variables.

$$p(\mathcal{D}_{\text{fut}}, \mathcal{D}_{\text{obs}}, \mathbf{z}) = \prod_i p(\mathcal{D}_i|\mathbf{z})p(\mathbf{z}) \tag{15}$$

- May not have a direct physical basis initially, but can turn out to be interpretable after training.

Latent Variables

How do we create a joint with interesting relationships between the observed and future data?

- Invent **latent variables** that are **common** to all data.

$$p(\mathcal{D}_{\text{fut}}, \mathcal{D}_{\text{obs}}, \mathbf{z}) \tag{14}$$

- For simplicity, often data is iid given latent variables.

$$p(\mathcal{D}_{\text{fut}}, \mathcal{D}_{\text{obs}}, \mathbf{z}) = \prod_i p(\mathcal{D}_i|\mathbf{z})p(\mathbf{z}) \tag{15}$$

- May not have a direct physical basis initially, but can turn out to be interpretable after training.
- Induces correlations between data, that can help to predict

$$p(\mathcal{D}_{\text{fut}}, \mathcal{D}_{\text{obs}}) = \int p(\mathcal{D}_{\text{fut}}, \mathcal{D}_{\text{obs}}, \mathbf{z})\mathrm{d}\mathbf{z} \tag{16}$$

Example: Linear Basis Function Regression

Linear regression falls under this!

$$p(\mathbf{y}, \mathbf{y}^*) = \int \prod_i p(y_i|\boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} \tag{17}$$

$$p(y_i|\boldsymbol{\theta}) = \mathcal{N}\left(y_i; \boldsymbol{\theta}^\top \boldsymbol{\phi}(\mathbf{x}), \sigma^2\right) \tag{18}$$

Example: Linear Basis Function Regression

Linear regression falls under this!

$$p(\mathbf{y}, \mathbf{y}^*) = \int \prod_i p(y_i|\boldsymbol{\theta}) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \tag{17}$$

$$p(y_i|\boldsymbol{\theta}) = \mathcal{N}\left(y_i; \boldsymbol{\theta}^\top \boldsymbol{\phi}(\mathbf{x}), \sigma^2\right) \tag{18}$$

- Do we really believe that the data we obtained was generated by sampling some parameters?
- Do we really believe that the relationship is a sum of polynomials?
- No, but it's useful for predicting!

Example: Linear Basis Function Regression

Linear regression falls under this!

$$p(\mathbf{y}, \mathbf{y}^*) = \int \prod_i p(y_i|\boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} \tag{17}$$

$$p(y_i|\boldsymbol{\theta}) = \mathcal{N}\left(y_i; \boldsymbol{\theta}^\top \boldsymbol{\phi}(\mathbf{x}), \sigma^2\right) \tag{18}$$

- Do we really believe that the data we obtained was generated by sampling some parameters?
- Do we really believe that the relationship is a sum of polynomials?
- No, but it's useful for predicting!

Similar, VAE (which we will also discuss).

How do we know we have a good model?

Model criticism is crucial!
Luckily, we have an objective metric on how well it’s doing:

How do we know we have a good model?

Model criticism is crucial!
Luckily, we have an objective metric on how well it's doing:

Predictive accuracy!

How do we know we have a good model?

Model criticism is crucial!
Luckily, we have an objective metric on how well it's doing:

Predictive accuracy!

- Hold-out test set.
- Check where it is overconfident and underconfident.
- Does it predict well when you change the setting?
- Bayesian model selection (soon).

How do we know we have a good model?

Model criticism is crucial!
Luckily, we have an objective metric on how well it's doing:

Predictive accuracy!

- Hold-out test set.
- Check where it is overconfident and underconfident.
- Does it predict well when you change the setting?
- Bayesian model selection (soon).

The ML philosophy: if you predict well, you understand.

Conclusion

Summary:

- You can do anything with the joint.
- Can create joints from understanding of the world.
- Can create joints by just hypothesising relationships.

Conclusion

Summary:

- You can do anything with the joint.
- Can create joints from understanding of the world.
- Can create joints by just hypothesising relationships.
  Just make sure you validate your model...

Conclusion

Summary:

- You can do anything with the joint.
- Can create joints from understanding of the world.
- Can create joints by just hypothesising relationships.
  Just make sure you validate your model...

What you should be able to do:

- Create probabilistic model (i.e. joints) by composing conditionals.
- Apply sum and product rules to find desired posteriors.

Conclusion

Summary:

- You can do anything with the joint.
- Can create joints from understanding of the world.
- Can create joints by just hypothesising relationships.
  Just make sure you validate your model...

What you should be able to do:

- Create probabilistic model (i.e. joints) by composing conditionals.
- Apply sum and product rules to find desired posteriors.

Reading & exercises:

- Chapter 3 (MacKay, 2003).
- Exercise: the burglar alarm (MacKay, 2003, ch.21)
- Exercise: bent coin (MacKay, 2003, §3.2)
- Exercise: legal evidence (MacKay, 2003, §3.4)

References I

MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.