Generalised additive models

## Transcript

Video 1: <https://youtu.be/A5FCExzZJm0>

Hello and welcome to this video introducing generalised additive models. Generalised additive models are statistical models that contain smooth functions of covariates allowing for non-linear relationships between our outcome of interest and one or more covariates. They offer a flexible alternative to the more traditional generalised linear modelling approaches that most people are familiar with. Generalised additive models are especially useful where we wish to explore non-linear relationship between our variables.

 Onscreen is the model equation for a generalised additive model, and if you’re familiar with linear models the first half of this should be fairly recognisable to you. That’s because the majority of this equation is exactly the same as our generalised linear modelling equation. On the left-hand side we have our outcome of interest, mu, transformed by some link function determined by model choice. For example, if we’re using a normal model then this link function would just be the identity function, so the left-hand side would simply be mu, whereas if we chose to use a Poisson model then this link function would be the log function.

 On the right-hand side the first element is alpha zero, which is our intercept value or can be thought of as a baseline term. This is the expected outcome where all of the values in our model take zero. Next we have a linear combination of some covariates, X, and coefficients estimated from the model. These assume a linear relationship between each of the covariates and the outcome of interest after adjusting for everything else in the model. And these coefficients give us an estimate of that linear relationship.

 The final term is what sets these generalised additive models apart. This is smooth functions applied to some covariates within our model. These smooth functions make our modelling framework far more flexible, but that flexibility does come at a cost. The inclusion of smooth functions require some specification of how that relationship looks, so we have to determine exactly what function we can apply to our data. We also have to determine how smooth or wiggly these functions must be, and it’s particularly important to do this because if we have an overly wiggly function, this could lead to us overfitting the data and could really limit the types of inferences that we can make about the population of interest.

 To address these concerns, generalised additive models use smoothing splines to construct the smooth functions. These splines are actually broken down further into linear combinations of basis functions applied at intervals across our coefficient and some coefficients estimated from the data. So, our smooth functions actually look like this, with the basis functions, b, determined before the model fit, and beta as estimated from the data.

 Basis functions are chosen based on before the model is fitted based on the type of smoothing spline we want to use. Each smoothing spline has a different basis function, so our choice of smoothing splint depends on the expected complexity and nature of the relationship we’re trying to capture and whether our smooth function is across a single variable or whether we want to generate this across multiple dimensions. Coefficients are estimated similarly to other regression coefficients. They’re based on the data and they actually act as weights for our basis functions. They determine the shape of this smooth.

 For example, if we’re expecting to have a fairly simple relationship between a single variable and our outcome, we may consider a cubic smoothing spline. So, the basis function goes along with this smoothing spline is the cubic function. This plot here shows unweighted cubic functions applied across intervals of our covariate X, so this is what the basis function looks like before it has been weighted by the coefficients.

 The animation onscreen shows how different combinations of the cubic basis function that we saw earlier, and different estimates of the coefficients can change the shape of the curve. As you can see, the nature of this relationship completely changes across different combinations of coefficients. That means that I could use the exact same basis function across many different situations and different sets of data but produce completely different relationships. So, how do we, or rather how does the software that we’re using estimate these coefficients?

 Well, there are many different ways that we could calculate these coefficients, but the general consensus across GAM experts is to use penalised log-likelihood, which produces the most stable results. This involves assigning a penalty to the log-likelihood of our model, and that ensures that the model does not overfit the data. Without this penalty then most likely what would happen is we would end up with a model that just interpolates our data, so it overfits the data, limiting the inferences we can make.

 The penalty consists of two terms, lambda and W. Lambda is known as our penalty parameter. If lambda gets larger and larger and tends towards infinity, that’s actually going to take our smooth function towards a straight line, whereas the closer to zero this lambda gets, the smaller the penalty will be and the more wiggly it becomes so it will tend towards this kind of interpolated piecewise function. So, this lambda is estimated as part of the model fitting process, and the model will aim to find some kind of middle ground between these two kind of extremes. The second component, W, is actually coming from our model. It’s a measure of model wiggliness. It’s based on the integral of the second derivative of the smooth function squared.

 Alongside the basis function and coefficients, another element that controls the shape of our smooth function is the number of knots or turning points in our basis functions. So, the larger the number of knots or turning points, the more wiggly this function is going to be. And if we go back to that original kind of determination of the smooth function, breaking it down into the linear combination of basis functions and coefficients, the number of basis functions and coefficients is determined by this K, so the higher the K is, the more turning points we have, the wigglier this smooth function is going to become.

 These three plots show how the number of knots changes the smoothness or wiggliness of our smooth function. On the left-hand side the first plot shows where we just have three knots or three basis functions, and you can see that doesn’t really capture the relationship between X and Y, it’s too smooth to pick up those peaks and troughs, whereas on the right-hand side K is set to 12 and that’s just a bit too wiggly, it’s overfitting the data, it’s picking up changes in the data that don’t necessarily reflect the overall patterns, whereas the middle plot is kind of what we’re aiming for, enough knots that it captures the nature of the relationship without totally overfitting and just explaining the sample that we have.

 In reality, we don’t actually specify K, it’s estimated alongside model fit. In reality, what we need to do as users of our software is to specify the maximum number of knots that we would allow in our model in order for it to not become over-wiggly. And what happens is that once we hit the correct maximum number, once it’s high enough, increasing it further is not actually going to improve the function and is not going to change the estimated number of knots. What it will do instead is rapidly increase the computational cost of fitting this model. Therefore what our aim is going to be when fitting these models is specifying a maximum number of knots which is high enough to capture the complexity of the model without slowing down the model fit too much. And this will be demonstrated further in video 3 when we check whether our K, our knots, has been set high enough using the software R.

 Thank you for listening. I hope that this has been a useful introduction to generalised additive models. The following videos are going to show an application of a generalised.

National Centre for Research Methods (NCRM)
Social Sciences
Murray Building (Bldg 58)
University of Southampton
Southampton SO17 1BJ
United Kingdom

**Web** www.ncrm.ac.uk
**Email** info@ncrm.ac.uk
**Tel** +44 23 8059 4539
**Twitter** @NCRMUK