Вопрос 1:
Почему так настойчиво говорят, что лагранжиан может зависеть только от времени, координат и их производных по времени, но не от импульсов, например?
Есть функция Лагранжа, а есть лагранжиан, и их надо различать.
Лагранжиан – это обычная физвеличина. И мы, в частности, можем написать

$$L = \frac{p^2}{2} - U(x)$$

(только p и x здесь именно обычные координаты, а не обобщённые, иначе эта формула неверна).
А функция Лагранжа – это именно что функция многих переменных $L(q, \dot{q}, t)$.

Говорят, что лагранжиан удобнее записывать как функцию Лагранжа (т.е. через $q, \dot{q}, t$ и только их) для дальнейшего решения уравнений Эйлера-Лагранжа. Но если вам вдруг решать их не надо (что маловероятно) – то можно записать как угодно.
Но, как правило, лагранжиан пишут для дальнейшего нахождения закона движения, поэтому сразу стоит его писать в каноническом виде, т.е. в виде функции Лагранжа.

Вопрос 2:
А может ли лагранжиан зависеть от вторых производных?
Процедура записи (то бишь придумывания лагранжиана) неформализована (в отличие от нахождения закона движения). Если вы поспрашиваете всяких теоретиков, они вам скажут, что они только и делают, что пытаются ПОДОБРАТЬ, что закон движения (вообще говоря, у них уже не закон движения, а поле) будет совпадать с экспериментом.

Написать вы можете, без проблем. Правда, «дефолтное» уравнение Эйлера-Лагранжа уже работать не будет, его придётся модернизировать. Так, если в функции Лагранжа будут производные вплоть до n-ной, то

$$\frac{\partial F}{\partial f} - \frac{d}{dx}\frac{\partial F}{\partial f'} + \frac{d^2}{dx^2}\frac{\partial F}{\partial f''} - \dots + (-1)^n \frac{d^n}{dx^n}\frac{\partial F}{\partial f^{(n)}} = 0$$

Но это чисто вычислительные сложности, а в остальном всё будет также.

То же касается производных не по координате. Скажем, вы можете написать

$$L = \frac{m}{2}\left(\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2\right) = \frac{m}{2}\left(\frac{dx}{dt}\right)^2\left(\left(\frac{dy}{dx}\right)^2 + 1\right)$$

Пожалуйста! Возвращаясь к первому вопросу: я записал таким образом ЛАГРАНЖИАН, а чтобы записать уравнения Эйлера-Лангража, таки придётся вернуться назад к «нормальной» форме, т.е. через функцию Лагранжа.

Вопрос 3:
Лагранжиан – это по определению L = T- U?
Думаю, вы уже поняли из предыдущих вопросов, то нет.
Вообще теоретики ответят на этот вопрос так:
1) Лагранжиан фундаментален и не обсуждается
2) Кинетической энергией называется слагаемое, где торчат квадраты первых производных (если оно есть, конечно) (т.е. он ВТОРИЧЕН, а Лагранжиан ПЕРВИЧЕН – естественно, с точки зрения теоретиков, а не «нормальных» людей-практиков)
3) Потенциальной энергией называется слагаемое, зависящей только от координат (если оно есть, конечно)
Таким образом, в лагранжиане могут быть оба слагаемых, может быть только одно, может быть что-то ещё (как правило, это «что-то ещё» обусловлено магнитным полем), может быть вообще

$$L = aq\dot{q}$$

Только вот будет ли реальная физсистема с таким лагранжианом? Но это уже не важно

На семинарах у вас будет стопицот задач, где вы будете лагранжиан ВЫПИСЫВАТЬ по алгоритму:
1) Считаем T
2) Считаем U
3) Вычитаем одно из другого
Но важно помнить, что всё это только с ИЛЛЮСТРАТИВНОЙ целью (причём вреда скорее больше, т.к. это и рождает стереотип «лагранжиан по определению равен T-U»). В серьёзной теорфизике люди как раз придумывают лагранжиан без всяких там T и U (как правило, на основании симметрий, теории групп, всякого такого).

Вопрос 4:
Мы вот пишем $L\left(t, q, \frac{dq}{dt}\right)$. Но если мы знаем функцию $q(t)$, мы знаем и её прозводную $\frac{dq}{dt}$. Почему мы не пишем просто $F(q, t)$?

Думаю, что часть читателей об этом и не задумывалась – а вот часть задумывался (я в своё время, например).

Этот вопрос ещё всплывёт на вариационном исчислении и, я, пожалуй, вставлю кусок оттуда сюда (там другие обозначения переменных, а в остальном то же):

Мы вот пишем $F\left(x, y, \frac{dy}{dx}\right)$. Но если мы знаем функцию $y(x)$, мы знаем и её прозводную $\frac{dy}{dx}$. Почему мы не пишем просто $F(x, y)$?

Давайте посмотрим на запись вида $F(y(x), y^2(x))$. Пусть x – это какое-то число, например, a. Тогда $F(y(a), y^2(a))$ – тоже число. Можем ли мы, зная первый аргумент - $y(a)$ в точке a, вычислить второй - $y^2(a)$? Да, надо просто возвести второй аргумент в квадрат. Т.е. вместо внешней функции **двух** переменных можем сделать обойтись функцией **одной** переменной:

$$F(y(x), y^2(x)) = g\big(y(x)\big)$$

где $g(z)$ по определению $= F(z, z^2)$

А теперь $F\left(y(x), \frac{dy}{dx}(x)\right)$. Мы знаем $y(a)$ в точке, можем ли вычислить $\frac{dy}{dx}(a)$? Нет – для этого нам нужно хоть немного, но окрестность точки a – одного значения функции в точке a нам не хватит! В этом и отличие.
Мы можем соорудить $\hat{g}(z) = F\left(z, \frac{dz}{dx}\right)$. Тогда $F\left(y(x), \frac{dy}{dx}(x)\right) = \hat{g}[y(x)]$. Но только вот $\hat{g}$ будет уже оператором (я не поленился и поставить крышку) – хотя бы потому, что само дифференцирование $\frac{dz}{dx}$ само по себе является оператором. А оператор – это «новый уровень», как говорится.
И обратите ещё внимание на квадратные скобки - $\hat{g}[y(x)]$: они подчёркивают, что оператору нужна вся функция y(x), а не только её значение в какой-то точке.

Упражнение. Сообразите, где можно сократить число переменных, т.е. заменить F(,) на функцию одной переменной:

$$1)\ F\left(y(x), \int_a^x y(t)dt\right)$$

$$2)\ F(y(x), y(x) - y(0))$$

$$3)\ F\left(\frac{dy(x)}{dx}, x - 1\right)$$

Ответ: только во 2).